What makes us different: Understanding the differences between human and machine-generated text

Διπλωματική Εργασία 1

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Explainability / Large Language Models
Τύπος Εργασίας	Research

Περιγραφή

In the digital age, where information flows freely and copiously, the distinction between human and machine intelligence becomes not just a technological challenge but also a societal one. As Large Language Models (LLMs) produce increasingly fluent text, the lines between human creativity and algorithmic efficiency blur. The capability to detect and understand the nuances differentiating human and machine outputs is essential, not only for maintaining trust and transparency in digital communication but also for comprehending the strengths and limitations of artificial intelligence. Furthermore, as machine-generated content finds its way into various domains—news, literature, law—it becomes paramount to investigate the fine line of differentiation and understand what makes human creation uniquely human.

Description of Work:

1. **Literature Review:** Begin with a comprehensive review of existing literature on text generation, focusing particularly on the distinctions between human and machine-generated content. This will cover methodologies currently employed in discerning textual origins, existing explainability techniques in the NLP realm, and human-centric studies on machine-generated content perception.
2. **ML/DL Method Exploration:** Explore current machine learning and deep learning techniques that can distinguish between human and machine-generated text. Assess their efficacy, strengths, and shortcomings.
3. **Explainability Focus:** Implement and analyze explainability techniques to unveil the underlying features and textual constructs that these models rely upon. This will shed light on what differentiates human linguistic patterns from algorithmically constructed ones. We will mainly experiment with counterfactual explanations, but we will investigate other types of explanations too.
4. **Human Evaluation and Perception:** Design and execute user studies to understand human perceptions and abilities in identifying machine-generated content. Explore intuitive as well as learned patterns in human discernment.
5. **Enhancing Human Judgment with Explanations:** Analyze whether machine-driven explanations and insights can assist humans in better identifying machine-generated content, effectively bridging the human-machine understanding gap.

6. **Algorithmic Variation:** Consider the differential impacts of texts generated by various models, ranging from smaller pretrained models to state-of-the-art LLMs, offering a holistic view of machine text generation capabilities.

Datasets and Tasks:

1. AuTexTification: Automated Text Identification shared task
2. Data Augmentation/Generation: Augment existing datasets by generating content using different LLMs and other pretrained language models, creating a rich ground for analysis.

Related Literature:

[1] Dugan, Liam, et al. "RoFT: A tool for evaluating human detection of machine-generated text." arXiv preprint arXiv:2010.03070 (2020).

[2] Jawahar, Ganesh, Muhammad Abdul-Mageed, and Laks VS Lakshmanan. "Automatic detection of machine generated text: A critical survey." arXiv preprint arXiv:2011.01314 (2020).

[3] Dugan, Liam, et al. "Real or fake text?: Investigating human ability to detect boundaries between human-written and machine-generated text." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. No. 11. 2023.

[4] Ross, Alexis, Ana Marasović, and Matthew E. Peters. "Explaining NLP models via minimal contrastive editing (MiCE)." arXiv preprint arXiv:2012.13985 (2020).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, και μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και τον Ορφέα Μενή - Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr)

Η Απόλυτη Μετριότητα:
Σύνθεση αντιπαραδειγμάτων μεγιστοποίησης της αβεβαιότητας ταξινόμησης

Διπλωματική Εργασία 2

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Νευρωνικά Δίκτυα και Βαθιά Μάθηση Διαθέσιμη
Status	Explainability
Περιοχή	Research
Τύπος Εργασίας	

Περιγραφή

“Το να φτάσεις την τελειότητα είναι δύσκολο, το να πετύχεις την απόλυτη μετριότητα είναι σχεδόν ακατόρθωτο.” - Albert Einstein

In the rapidly advancing domain of Natural Language Processing (NLP), interpretability and explainability of models have become paramount. Conventional counterfactual explanations in NLP have mainly focused on providing clear distinctions between classes by drastically shifting classification probabilities. Yet, understanding the nuanced 'semantic borders'—the regions where the model is uncertain—can offer deeper insights into the model's decision-making process. When deploying machine learning models in critical environments, simply knowing when the model is "wrong" (via traditional counterfactuals) isn't enough. Evaluators might want to explore the model's behavior in gray zones to gauge its reliability. For instance, in automated medical diagnostics, an evaluator might want to know not just when a model changes its diagnosis, but how it behaves when a diagnosis isn't clear-cut. Here, SBCs can serve as stress tests, helping evaluators identify situations where the model's judgment is potentially unreliable. This can not only aid in comprehending the intricacies of the classifier's inner workings but also unveil potential vulnerabilities and biases. Additionally, generating counterfactual examples on this border can provide invaluable synthetic data, enhancing the robustness of the model and possibly revealing ambiguous or poorly defined class boundaries. Such a focus also aligns with the real-world scenarios where many decisions are not black and white but rather dwell in shades of grey. Capturing this uncertainty aids in creating models that are more aligned with human-like decision-making processes.

Description of Work:

1. **Literature Review:** Begin by reviewing the existing literature on counterfactual explanations in NLP, focusing on both traditional methods and the more recent advancements. This will provide the necessary foundation for understanding the current state-of-the-art and identifying gaps in the existing approaches.
2. **Calibrated Classifier Integration:** Incorporate calibrated classifiers to ensure that the probabilities output by the model genuinely represent the likelihood of each class. This ensures that the identified

semantic borders/decision boundaries are based on genuine model uncertainty rather than mis-calibrated scores.

3. **Development of Semantic Border Counterfactuals (SBC):** Design and implement an algorithm or method for generating counterfactual explanations that lie on the semantic border. This will involve modifying inputs such that their classification probabilities hover around 0.5, indicating a region of uncertainty.
4. **Evaluation:** Quantitatively and qualitatively evaluate the generated counterfactuals in terms of their relevance, pertinence, and clarity. This will involve establishing metrics to measure the quality of explanations and potentially comparing them to traditional counterfactual methods. This may include user surveys on evaluating the effect of SBCs on understanding the classifier.
5. **Synthetic Data Generation:** Explore the feasibility of using the generated SBCs as synthetic data to augment training datasets, assessing their utility in improving model robustness and generalization.

Datasets and Tasks:

1. **Sentiment Analysis**: Using datasets like the IMDB Movie Reviews or the Stanford Sentiment Treebank, evaluate how models handle ambiguous sentiments that lie on the boundary between positive and negative.
2. **Topic Classification:** Datasets such as the 20 Newsgroups, the 7 Newsgroups, or the Reuters news dataset can be employed to analyze counterfactuals in the context of topics that may span multiple categories or lie on the border between them.
3. **Toxic Comment Classification:** Utilizing the Jigsaw Toxic Comment Classification Challenge dataset, assess how models discern comments that teeter on the edge of being toxic or non-toxic.

Throughout these tasks, emphasis will be placed on calibrated classifiers such as Platt Scaling or Isotonic Regression, ensuring genuine interpretability in the uncertainty regions of the models.

Related Literature:

[1] Ross, Alexis, Ana Marasović, and Matthew E. Peters. "Explaining NLP models via minimal contrastive editing (MiCE)." arXiv preprint arXiv:2012.13985 (2020).
[2] Jin, Di, et al. "Is bert really robust? a strong baseline for natural language attack on text classification and entailment." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 05. 2020.
[3] Filandrianos, Giorgos, et al. "Counterfactuals of Counterfactuals: a back-translation-inspired approach to analyse counterfactual editors." arXiv preprint arXiv:2305.17055 (2023).
[4] Ross, Alexis, et al. "Tailor: Generating and perturbing text with semantic controls." arXiv preprint arXiv:2107.07150 (2021).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, μέθοδοι ερμηνείας μοντέλων μηχανικής μάθησης και αντιπαραθετικων παραδειγματων (counterfactual examples). Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και τον Ορφέα Μενή - Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

BrainTeaser: commonsense question-answering for lateral thinking using Large Language Models

Διπλωματική Εργασία 3

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
	Μη Διαθέσιμη
Status	Large Language Models
Περιοχή	Research
Τύπος Εργασίας	

Περιγραφή

Human reasoning processes comprise two types of thinking: vertical and lateral. Vertical thinking is a sequential analytical process that is based on rationality, logic, and rules. Meanwhile, lateral thinking (or “thinking outside the box”) is a divergent and creative process that involves looking at a problem from a new perspective and defying preconceptions.

The success of language models has inspired the NLP community to attend to tasks that require implicit and complex reasoning, relying on human-like commonsense mechanisms. While such vertical thinking tasks have been relatively popular, lateral thinking puzzles have received little attention. To bridge this gap, we devise BRAINTEASER: a multiple-choice Question Answering task designed to test the model’s ability to exhibit lateral thinking and defy default commonsense associations.

BRAINTEASER QA task consists of two subtasks-Sentence Puzzle and Word Puzzle that require awareness of commonsense “defaults” and overwriting them through unconventional thinking that distinguishes these defaults from hard constraints.

Sentence Puzzle: Sentence-type brain teaser where the puzzle defying commonsense is centered on sentence snippets. For example:
Q:
A man shaves everyday, yet keeps his beard long.
A candidates:
He is a barber. ✅
He wants to maintain his appearance.
He wants his girlfriend to buy him a razor.
None of the above.

Word Puzzle: Word-type brain teaser where the answer violates the default meaning of the word and focuses on the letter composition of the target question. For example:
Q:

What part of London is in France?
A candidates:
The letter N. ✅
The letter O.
The letter L.
None of the above.

Both tasks include an adversarial subset, created by manually modifying the original brain teasers without changing their latent reasoning path.
For more information: https://semevalbrainteaser.github.io and sample data https://drive.google.com/drive/u/2/folders/1kiFXp5fqpf8--NQJJAlBIBpfSaXTk1UY

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr) και το Γιώργο Φιλανδριανό (geofila@islab.ntua.gr)

Multigenerator, Multidomain, and Multilingual Black-Box Machine-Generated Text Detection

Διπλωματική Εργασία 4

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
	Μη Διαθέσιμη
Status	Large Language Models
Περιοχή	Research
Τύπος Εργασίας	

Περιγραφή

Large language models (LLMs) are becoming mainstream and easily accessible, ushering in an explosion of machine-generated content over various channels, such as news, social media, question-answering forums, educational, and even academic contexts. Recent LLMs, such as ChatGPT and GPT-4, generate remarkably fluent responses to a wide variety of user queries. The articulate nature of such generated text makes LLMs attractive for replacing human labor in many scenarios. However, this has also resulted in concerns regarding their potential misuse, such as spreading misinformation and causing disruptions in the education system. Since humans perform only slightly better than chance when classifying machine-generated vs. human-written text, there is a need to develop automatic systems to identify machine-generated text with the goal of mitigating its potential misuse. We offer three subtasks over two paradigms of text generation: (1) full text completely written by humans or generated by a machine; and (2) mixed text (machine-generated text refined by human or human-written text paraphrased by a machine).

Subtask A. Binary Human-Written vs. Machine-Generated Text Classification: Given a full text, determine whether it is human-written or machine-generated.

Subtask B. Multi-Way Machine-Generated Text Classification: Given a full text, determine who generated it. It can be human-written or generated by a specific language model.

Subtask C. Human-Machine Mixed Text Detection: Given a mixed text, where the first part is human-written and the second part is machine-generated, determine the boundary, where the change occurs.

More info here: https://github.com/mbzuai-nlp/SemEval2024-task8

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr) και τον Ορφέα Μενή - Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr).

Numeral-Aware Language Understanding and Generation

Διπλωματική Εργασία 5

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση Διαθέσιμη
Status	Large Language Models
Περιοχή	Research
Τύπος Εργασίας	

Περιγραφή

Despite the recent success of Large Language Models, understanding and generation of numerical values has been underestimated in comparison to the attention given to the semantic and grammatical quality achieved in purely non-numerical textual examples. However, numbers play a significant role in related corpora: Consider, for instance, a scenario where one anticipates a 30% rise in stock prices versus a 3% rise. This nuance plays a pivotal role in fine-grained sentiment analysis, as the former implies a stronger sentiment than the latter. Similarly, in a legal context, the statement "Stealing 10 dollars" compared to "Stealing 100,000 dollars" could potentially lead to differing court judgments.

The problem of numerical understanding and generation can be divided into three subtasks, as described here: https://sites.google.com/view/numeval/tasks?authuser=0 . We will focus our efforts on English datasets. Data samples for all subtasks are provided here: https://sites.google.com/view/numeval/data?authuser=0

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr) και τον Γιώργο Φιλανδριανό (E-mail: geofila@islab.ntua.gr).

SHROOM, a Shared-task on Hallucinations and Related Observable Overgeneration Mistakes

Διπλωματική Εργασία 6

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
	Διαθέσιμη
Status	Large Language Models
Περιοχή	Research
Τύπος Εργασίας	

Περιγραφή

The modern NLG landscape is plagued by two interlinked problems: On the one hand, our current neural models have a propensity to produce inaccurate but fluent outputs; on the other hand, our metrics are most apt at describing fluency, rather than correctness. This leads neural networks to “hallucinate”, i.e., produce fluent but incorrect outputs that we currently struggle to detect automatically. For many NLG applications, the correctness of an output is however mission critical. For instance, producing a plausible-sounding translation that is inconsistent with the source text puts in jeopardy the usefulness of a machine translation pipeline. With our shared task, we hope to foster the growing interest in this topic in the community.

With SHROOM we adopt a post hoc setting, where models have already been trained and outputs already produced: in this research thesis, the student will be asked to perform binary classification to identify cases of fluent overgeneration hallucinations in two different setups: model-aware and model-agnostic tracks. That is, the student must detect grammatically sound outputs which contain incorrect or unsupported semantic information, inconsistent with the source input, with or without having access to the model that produced the output. To that end, we will provide the student with a collection of checkpoints, inputs, references and outputs of systems covering four different NLG tasks: definition modeling (DM), machine translation (MT), paraphrase generation (PG) and text simplification (TS), trained with varying degrees of accuracy. The development set will provide binary annotations from at least five different annotators and a majority vote gold label.

For more information and data check here: https://helsinki-nlp.github.io/shroom/

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr) και τον Γιώργο Φιλανδριανό (E-mail: geofila@islab.ntua.gr).

Explaining vision-language models via linguistic perturbations

Διπλωματική Εργασία 7

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
	Διαθέσιμη
Status	Large Language Models, Explainability
Περιοχή	Research
Τύπος Εργασίας	

Περιγραφή

Vision-Language (VL) learning has been one of the most popular fields in deep learning, counting numerous successful approaches that tackle several tasks. Apart from performance improvements, explainability has been one of the major challenges accompanying the development of state-of-the-art VL models. Some prior endeavors such as those analyzed in our paper 'Knowledge-Based Counterfactual Queries for Visual Question Answering' https://arxiv.org/pdf/2303.02601.pdf have demonstrated some interesting insights regarding the internal workings of Visual Question Answering (VQA) tasks.

In the current thesis, we plan to expand the experimentation conducted in the aforementioned paper to cover more models and datasets. Specifically, we can replicate the techniques on tasks such as Visual Commonsense reasoning (VCR) and Visual Word Sense Disambiguation (VWSD) to evaluate the sensitivity of related models. Moreover, we can craft additional knowledge-based linguistic attacks to obtain a larger set of insights from these models and measure their sensitivity to linguistic perturbations.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr),τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr) και τον Γιώργο Φιλανδριανό (E-mail: geofila@islab.ntua.gr)

Unbiased scene graph generation via Large Language Models

Διπλωματική Εργασία 8

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Large Language Models, Explainability
Τύπος Εργασίας	Research

Περιγραφή

Scene graph generation (SGG) is an ongoing research problem which encounters several issues, such as missing or erroneous objects and relationships. This is due to the fact that most SGG models are trained on specific datasets such as Visual Genome, which contain many imperfect annotations, leading to learning errors and biases. These errors are consistently depicted on generated graphs and impact downstream tasks, such as graph-based explanations.

In this thesis, we propose alternative methods of producing scene graphs to try and tackle related issues, leveraging the knowledge hidden in models such as LLMs. For example, we can produce captions from images and then extract triples from the text. Moreover, in the multimodal setting we can immediately produce triples from the given images. Since this task has not been explored in recent literature, the student needs to explore several prompting techniques to achieve retrieving the appropriate knowledge. Therefore, a lot of effort needs to be invested in order to deliver this specific thesis.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr), Γιώργος Φιλανδριανός (geofila@islab.ntua.gr) και Κωνσταντίνος Θωμάς (kthomas@islab.ntua.gr).

Bridging Semantic Counterfactuals and Image Synthesis

Διπλωματική Εργασία 9

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Explainability, Image Generation
Τύπος Εργασίας	

Περιγραφή

Developing an algorithm for generating counterfactual images from graphs, based on a source image, represents an innovative and complex research endeavor. This task entails overcoming inherent challenges associated with semantic counterfactual algorithms, which primarily operate on graphs to produce edits, and expanding their capabilities to craft entirely new counterfactual images. Unlike traditional graph-based counterfactual algorithms, this novel approach necessitates the transformation of graph-based edits into visually coherent and semantically meaningful images.

One of the primary objectives of this research is to harness the insights and capabilities offered by semantic counterfactual algorithms that work with graph structures. These algorithms typically excel in identifying and manipulating relationships and attributes within graph data. In this context, the challenge is to leverage these algorithmic insights to create counterfactual images that maintain a high degree of visual fidelity to the source image while altering specific aspects to reflect counterfactual scenarios.

Furthermore, this research extends the existing paradigm by bridging the gap between graph-based manipulations and image synthesis. Unlike traditional graph-to-graph transformations, the generation of counterfactual images from graphs remains an underexplored area in recent literature. This work requires the exploration of various techniques, possibly involving both graph-based and image-based machine learning models, to successfully achieve the transformation from graphs to visually coherent images.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Γιώργο Φιλανδριανό (geofila@islab.ntua.gr) και τον Κωνσταντίνο Θωμά (kthomas@islab.ntua.gr).

Ανάπτυξη συστήματος αυτόματης εναρμόνισης μουσικής μελωδίας

Διπλωματική Εργασία 10

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενο μάθημα	Τεχνητή Νοημοσύνη
Status	Διαθέσιμη
Περιοχή	
Τύπος Εργασίας	

Περιγραφή

Ένα μουσικό έργο συνήθως αναλύεται σε τρεις βασικές συνιστώσες: τη μελωδία, την αρμονία και τον ρυθμό. Κάθε δημιουργός επιλέγει τον τρόπο που θα συνθέσει το έργο του, είτε ξεκινώντας από τη μελωδία και στη συνέχεια την εναρμόνισή της, είτε ξεκινώντας δομώντας την αρμονία πάνω στην οποία θα βασίσει τη μελωδία. Σκοπός της συγκεκριμένης διπλωματικής είναι η μελέτη και η ανάπτυξη εργαλείων αυτόματης εναρμόνισης μιας δοθείσας μελωδίας. Για το σκοπό αυτό θα συλλεγούν μεγάλα δεδομένα μουσικών έργων σε συμβολική μορφή για την εξαγωγή πληροφορίας καθώς και θα αναπτυχθεί μια Βάση Γνώσης για την κατηγοριοποίηση βασικών κανόνων εναρμόνισης. Στη συνέχεια, θα αναπτυχθεί ένα σύστημα, το οποίο δοθείσης μιας μελωδίας θα μπορεί να προτείνει και να συγκρίνει διαφορετικούς τρόπους εναρμόνισής της καθώς και να αναπαράγει το ακουστικό αποτέλεσμα.

Καθώς η περιοχή της ανάκτησης μουσικής πληροφορίας συνδυάζει τεχνικές μηχανικής μάθησης, συστημάτων γνώσης, είναι σημαντικό ο/η φοιτητής/τρια που θα αναλάβει τη συγκεκριμένη εργασία να έχει γνώση και των δύο περιοχών. Ακόμη, επειδή θα μελετηθούν τεχνικές εναρμόνισης βασισμένες τόσο στην τονική αρμονία όσο και σε σύγχρονους τρόπους όπως η χρήση ακολουθιών συγχορδιών (chord progressions), είναι αναγκαία η εξοικείωση και η κατανόηση σε βάθος σύνθετων μουσικών όρων. Η εργασία έχει έντονο ερευνητικό αλλά και δημιουργικό χαρακτήρα, οπότε θα ήταν επιθυμητό ο/η φοιτητής/τρια να έχει μια προσωπική άποψη πάνω στη μουσική σύνθεση και να μπορεί να προτείνει δικούς/ές του/της τρόπους εναρμόνισης.

Απαιτούμενες/επιθυμητές γνώσεις: Περιγραφικές Λογικές, Γράφοι Γνώσης, Τεχνικές Αναπαράστασης Γνώσης, Οντολογίες, εξοικείωση με κάποια Γλώσσα Προγραμματισμού (ενδεικτικά Python), κατανόηση σε βάθος μουσικών όρων.Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr)

How to Go Viral: Leveraging Graph and Semantic Counterfactual Algorithms

Διπλωματική Εργασία 11

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενο μάθημα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	
Περιοχή	Διαθέσιμη
Τύπος Εργασίας	

Περιγραφή

The process of transforming ordinary YouTube videos into viral sensations remains a complex and evolving challenge. This research endeavor revolves around the creation of a specialized dataset comprising YouTube video thumbnails and corresponding textual metadata, specifically targeting a defined subsection of videos. The objective is to curate a dataset that encapsulates a variety of content characteristics, engagement metrics, and video attributes to facilitate a comprehensive analysis.

The pivotal aspect of this research lies in the transformation of these video-related data into graph representations. Graphs offer a structured and versatile framework for capturing the intricate relationships and dependencies inherent in video content, viewer interactions, and external factors affecting virality. This transformation involves defining nodes and edges that encode various video attributes, viewer behaviors, and contextual information. Subsequently, the research delves into the realm of graph counterfactual algorithms, where the focus is on deciphering the critical factors responsible for a video's transition from non-viral to viral status. Graph-based algorithms and semantic counterfactual techniques are employed to manipulate and analyze the constructed graph structures. This entails exploring hypothetical scenarios, perturbing graph elements, and identifying key interventions that lead to increased video virality.The primary goal of this research is to uncover the underlying principles and strategies that contribute to a video's success. It involves a deep investigation into the graph-based representations of videos, encompassing aspects such as content optimization, audience engagement, and external promotion. The exploration of counterfactual scenarios within the graph context aims to shed light on the specific actions or modifications required to enhance a video's potential for virality.

Given the complexity of this interdisciplinary task, substantial effort and a multifaceted approach are essential to deliver a comprehensive understanding of the factors influencing video success and to devise effective strategies for video content creators and platform optimization.

Απαιτούμενες/επιθυμητές γνώσεις: Περιγραφικές Λογικές, Γράφοι Γνώσης, Τεχνικές Αναπαράστασης Γνώσης, Οντολογίες, εξοικείωση με κάποια Γλώσσα Προγραμματισμού (ενδεικτικά Python), κατανόηση σε βάθος μουσικών όρων.Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Γιώργο Φιλανδριανό (geofila@islab.ntua.gr) και τον Κωνσταντίνο Θωμά (kthomas@islab.ntua.gr).

Ανάπτυξη μουσικού Γράφου Γνώσης για την αναγνώριση μοτίβων

Διπλωματική Εργασία 12

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενο μάθημα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	
Περιοχή	Διαθέσιμη
Τύπος Εργασίας	

Περιγραφή

Το πεδίο της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval - MIR) είναι ένας χώρος συνδεδεμένος με την Επιστήμη των Υπολογιστών. Υπάρχει μεγάλη πληθώρα δεδομένων τόσο σε ηχητική μορφή (audio) όσο και σε συμβολική (midi, MusicXML, txt). Σκοπός της συγκεκριμένης εργασίας είναι η ανάπτυξη ενός μουσικού Γράφου Γνώσης βασισμένο σε έναν μεγάλο όγκο δεδομένων συμβολικής μουσικής με στόχο την αναγνώριση μοτίβων βάσει των αρμονικών χαρακτηριστικών των μουσικών κομματιών.
Καθώς η περιοχή της ανάκτησης μουσικής πληροφορίας συνδυάζει τεχνικές μηχανικής μάθησης, συστημάτων γνώσης, είναι σημαντικό ο/η φοιτητής/τρια που θα αναλάβει τη συγκεκριμένη εργασία να έχει γνώση και των δύο περιοχών. Ακόμη, επειδή θα μελετηθούν μουσικά μοτίβα βασισμένα σε αρμονικά χαρακτηριστικά, είναι αναγκαία η εξοικείωση και η κατανόηση σε βάθος σύνθετων μουσικών όρων. Η εργασία έχει έντονο ερευνητικό αλλά και δημιουργικό χαρακτήρα, οπότε θα ήταν επιθυμητό ο/η φοιτητής/τρια να έχει μια προσωπική άποψη πάνω στη μουσική.

Απαιτούμενες/επιθυμητές γνώσεις: Περιγραφικές Λογικές, Γράφοι Γνώσης, Τεχνικές Αναπαράστασης Γνώσης, εξοικείωση με κάποια Γλώσσα Προγραμματισμού (ενδεικτικά Python), κατανόηση σε βάθος μουσικών όρων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr)

Ανάπτυξη Οντολογίας για την συσχέτιση μουσικής και συναισθήματος

Διπλωματική Εργασία 13

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενο μάθημα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	
Περιοχή	Διαθέσιμη
Τύπος Εργασίας	

Περιγραφή

Η τέχνη της μουσικής συνδέεται άμεσα με την πρόκληση συναισθημάτων στον ακροατή της. Από τη σύλληψη ενός μουσικού έργου μέχρι την τελική παραγωγή του και την ακρόασή του, ο δημιουργός προσπαθεί να μεταδώσει και να προκαλέσει συναισθήματα στον ακροατή. Υπάρχουν αρκετές τεχνικές οι οποίες μπορούν να χρησιμοποιηθούν από έναν δημιουργό γι’ αυτό τον σκοπό: επιλογή νοτών, τέμπο, ενορχήστρωσης, κ.ά. Κάθε μία από αυτές τις τεχνικές - και οι συνδυασμοί τους - μπορούν να συσχετιστούν με αντίστοιχα συναισθήματα.

Στη συγκεκριμένη εργασία ο/η φοιτητής/τρια θα αναπτύξει μια Οντολογία ώστε να συσχετίσει τις τεχνικές αυτές με αντίστοιχα συναισθήματα τα οποία μπορούν να προκαλέσουν. Θα μελετηθεί συγκεκριμένη βιβλιογραφία που αντιστοιχεί τεχνικές και συναισθήματα. Θα γίνουν παρατηρήσεις πάνω στην υπάρχουσα βιβλιογραφία και θα αναπτυχθούν σχέσεις και κανόνες μεταξύ τεχνικών και συναισθημάτων με την χρήση Περιγραφικών Λογικών καθώς και με τεχνολογίες του Σημασιολογικού Ιστού.

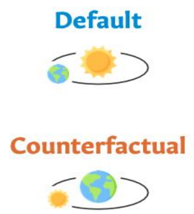
Απαιτούμενες/επιθυμητές γνώσεις: Περιγραφικές Λογικές, Οντολογίες, Τεχνικές Αναπαράστασης Γνώσης, εξοικείωση με κάποια Γλώσσα Προγραμματισμού (ενδεικτικά Python), κατανόηση μουσικών όρων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr)

Dataset Creation for Real-World and Counterfactual World Reasoning: Evaluating NLP Model's Reasoning Abilities

Διπλωματική Εργασία 14

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Dataset

Περιγραφή



The development of a dataset that facilitates reasoning within both real and imaginary contexts serves as a pivotal research endeavor. This task involves curating a comprehensive dataset specifically designed to challenge natural language processing (NLP) models in their ability to reason effectively. The dataset aims to evaluate the model's capacity for reasoning across a spectrum of scenarios, encompassing both real-world and imaginative[1] situations [1].

The dataset creation process is tailored to include a wide array of questions and prompts that require reasoning abilities. For instance, questions involving age comparisons, such as determining if a person who is 30 years old is older than someone who is 40 years old when considering reversed age, are incorporated. To ensure a robust evaluation, the dataset includes not only the questions but also the correct answers, serving as a benchmark for assessing the model's reasoning accuracy.

Given the novelty and complexity of this task, considerable effort is essential in constructing a dataset that effectively captures the nuances of reasoning in both real and imaginary worlds. The resulting dataset will serve as a valuable resource for assessing the reasoning abilities of NLP models, shedding light on their capacity to navigate the complexities of varied contexts.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr), Γιώργος Φιλανδριανός (geofila@islab.ntua.gr) και Κωνσταντίνος Θωμάς (kthomas@islab.ntua.gr).

[1] Wu, Zhaofeng, et al. "Reasoning or reciting? exploring the capabilities and limitations of language models through counterfactual tasks." *arXiv preprint arXiv:2307.02477* (2023).

[1] https://github.com/google/BIG-bench/tree/main/bigbench/benchmark_tasks/fantasy_reasoning

Hit Song Prediction by the trends *(Would the #1 hit song of 2013 be a hit if it was released in 2023?)*

Διπλωματική Εργασία 15

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενο μάθημα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	
Περιοχή	Διαθέσιμη
Τύπος Εργασίας	

Περιγραφή

Όλα τα μουσικά έργα μοιράζονται έναν κοινό στόχο: να απευθυνθούν σε όσο μεγαλύτερο κοινό μπορούν. Η επιτυχία, όμως, ενός μουσικού έργου εξαρτάται από πολλούς παράγοντες οι οποίοι μπορεί να είναι διακριτοί μεταξύ τους. Εκτός από τα μουσικά χαρακτηριστικά και τους στίχους ενός κομματιού, σημαντικό ρόλο για την επιτυχία του παίζουν ο καλλιτέχνης, η εποχή στην οποία κυκλοφορεί, αλλά και το marketing (πρόσφατο παράδειγμα ότι σχεδόν όλα τα τραγούδια της ταινίας Barbie βρέθηκαν στα top charts του Spotify). Μέχρι στιγμής, οι περισσότερες δουλειές πάνω στο αντικείμενο δίνουν έμφαση κυρίως στη μουσικά χαρακτηριστικά και στους στίχους, αγνοώντας τους υπόλοιπους παράγοντες.

Σκοπός της εργασίας είναι η συλλογή μεταδεδομένων για έναν μεγάλο όγκο μουσικών κομματιών και η ανάπτυξη ενός συστήματος μηχανικής μάθησης που θα προβλέπει την επιτυχία ενός μουσικού κομματιού, λαμβάνοντας υπόψη και συσχετίζοντας διαφορετικά και ανομοιογενή κριτήρια.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και Σπύρο Κανταρέλη (e-mail: spyroskanta at ails.ece.ntua.gr)

Audio Chord Estimation + Audio Key Detection

Διπλωματική Εργασία 16

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενο μάθημα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	
Τύπος Εργασίας	

Περιγραφή

Η εκτίμηση (estimation) μιας συγχορδίας και ο εντοπισμός της τονικότητας (key detection) ενός μουσικού κομματιού σε ηχητική μορφή (audio) είναι ένα ανοιχτό ζήτημα στο πεδίο της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval - MIR) (https://www.music-ir.org/mirex/wiki/2021:Audio_Chord_Estimation). Οι περισσότερες εργασίες που έχουν προταθεί βασίζονται σε μοντέλα βαθιάς μάθησης και παρουσιάζουν αρκετά καλά αποτελέσματα. Σκοπός της εργασίας είναι η μελέτη των μοντέλων αυτών και η προσπάθεια βελτίωσης της απόδοσής τους με την χρήση Τεχνικών Αναπαράστασης Γνώσης πάνω στα δεδομένα που θα αφορούν τα αρμονικά χαρακτηριστικά τους.
Καθώς η περιοχή της ανάκτησης μουσικής πληροφορίας συνδυάζει τεχνικές μηχανικής μάθησης και συστημάτων γνώσης, είναι σημαντικό ο/η φοιτητής/τρια που θα αναλάβει τη συγκεκριμένη εργασία να έχει γνώση και των δύο περιοχών. Ακόμη, επειδή θα μελετηθούν μουσικά μοτίβα βασισμένα σε αρμονικά χαρακτηριστικά, είναι αναγκαία η εξοικείωση και η κατανόηση σε βάθος σύνθετων μουσικών όρων

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων, Γράφοι Γνώσης, Τεχνικές Αναπαράστασης Γνώσης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και Σπύρο Καντaρέλη (e-mail: spyroskanta at ails.ece.ntua.gr)

What if: a commonsense reasoning counterfactual dataset

Διπλωματική Εργασία 17

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Dataset

Περιγραφή

Counterfactual reasoning is a cognitive process of thinking about and exploring hypothetical scenarios or events that did not actually happen but could have happened under different conditions; for example, what would have happened if I dropped a glass full of water?. It involves imagining alternative outcomes or courses of action by changing one or more factors while keeping other variables constant. Counterfactual reasoning is a fundamental aspect of human thinking and is used in various contexts, including decision-making, learning, and understanding causality.

Despite its significance, there is no systematic dataset containing counterfactual real-world statements expressing commonsense knowledge. This project aims to cover this gap by gathering 'what if' and 'what could have happened' statements addressing a variety of commonsense reasoning situations. Such statements could be used as prompts to Large Language Models (LLMs) to evaluate not only their reasoning capabilities in the factual setting, but also in the counterfactual one, evaluating to which extent LLMs are capable of counterfactual thinking in comparison to the factual thinking process. This dataset should also include counterfactual statements with negations (aka 'what if I did not…) since there is evidence that state-of-the-art LLMs struggle with negation. This dataset can be used in the future for training related models, enabling them to distinguish between real and hypothetical settings.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr) και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr), Γιώργος Φιλανδριανός (geofila@islab.ntua.gr) και Κωνσταντίνος Θωμάς (kthomas@islab.ntua.gr).

Explainable Decision Trees: Beyond Built-in Interpretability

Διπλωματική Εργασία 18

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Explainability
Τύπος Εργασίας	Research

Περιγραφή

Decision trees are widely used in machine learning because of their inherent transparency; their hierarchical structure can often be visualized and understood even by those without a technical background. However, this built-in interpretability is not always synonymous with real-world explainability. The structure, complexity, and terminology used in a decision tree might make sense to an algorithm, but can still be cryptic or counter-intuitive to human stakeholders. The primary aim of this thesis is to develop methodologies and frameworks to enhance the explainability of decision trees, by emphasizing their structure and vocabulary. The research will focus on ensuring that the decision trees generated are not only interpretable by design but also align with human cognitive processes and domain-specific terminologies.

Research Goals:

- **Analyze Existing Gaps:** Begin with an in-depth analysis of current decision tree algorithms to identify gaps in their explainability, especially when dealing with complex datasets or domain-specific applications.
- **Human-Centric Design:** Investigate cognitive science and human decision-making processes to inform the structure and design of more explainable decision trees.
- **Vocabulary Enhancement:** Develop a methodology to adapt the decision tree splits and node descriptions to employ domain-specific vocabulary, ensuring that the tree's decisions are communicated in terms that are immediately relevant and comprehensible to domain experts.
- **Complexity Management:** Propose techniques to manage the complexity of decision trees, balancing accuracy with simplicity, so that the trees remain both effective and understandable.
- **Evaluation Metrics:** Create new metrics or adapt existing ones to measure the explainability of decision trees, factoring in both their structure and vocabulary.
- **Case Studies:** Implement the developed methodologies on real-world datasets to showcase their effectiveness. Compare the generated trees with traditional trees in terms of both accuracy and explainability.

This research aims to push the boundaries of what it means for a decision tree to be "explainable." By redefining the structure and vocabulary of decision trees, the goal is to make them more transparent, understandable, and useful across a wider array of real-world applications. The outcomes could redefine best practices in decision tree generation and use, ensuring they remain a trusted tool in the age of complex machine learning models.

Related Literature:
[1] Bertsimas, Dimitris, and Jack Dunn. "Optimal classification trees." Machine Learning 106 (2017): 1039-1082.
[2] Jena, Monalisa, and Satchidananda Dehuri. "DecisionTree for Classification and Regression: A State-of-the Art Review." Informatica 44.4 (2020).
[3] Jimmy Lin, Chudi Zhong, Diane Hu, Cynthia Rudin, Margo Seltzer. “Generalized and Scalable Optimal Sparse Decision Trees.” Proceedings of the 37th International Conference on Machine Learning (2020)
[4] Costa, Vinícius G., and Carlos E. Pedreira. "Recent advances in decision trees: An updated survey." *Artificial Intelligence Review* 56.5 (2023): 4765-4800.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Ορφέα Μενή - Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr), και τον Ιάσονα Λιάρτη. (e-mai: jliartis at ails.ece.ntua.gr)

Hallucinations of Large Language Models

Διπλωματική Εργασία 19

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Survey

Περιγραφή

Large Language Models (LLMs) have undoubtedly solidified their position as a dominant paradigm in the realm of Natural Language Processing (NLP), boasting impressive language generation capabilities and wide-ranging applications. However, this meteoric rise in popularity has not been without its share of concerns, with one of the most prominent being the issue of hallucinations. Hallucinations, in the context of LLMs, refer to instances where these models produce responses that deviate from factual accuracy or introduce fabricated information into their output. These concerns have sparked rigorous investigation within the NLP community, seeking to unveil the intricate mechanisms behind LLM hallucinations and

The phenomenon of LLM hallucinations poses multifaceted challenges that extend beyond merely acknowledging their existence. Researchers are confronted with the daunting task of deciphering why and when these hallucinations occur. One prevailing question is whether LLMs produce erroneous responses when they lack access to the genuine answer to a given prompt or if these hallucinations emerge as a result of incomplete reasoning processes within the models. This conundrum underscores the complexity of LLM behavior, highlighting the need for a nuanced understanding of the intricate interplay between knowledge representation, reasoning capabilities, and the generation of human-like text. Addressing these questions is vital for advancing the field of NLP and harnessing the full potential

In pursuit of these answers, this literature review project embarks on a mission to construct a comprehensive taxonomy that categorizes and dissects the various sources and manifestations of LLM hallucinations. By conducting a thorough analysis of recent literature, the project seeks to cast a spotlight on the primary factors contributing to hallucinations and the nuanced contexts in which they arise. Through this meticulous survey paper, the student spearheading this research endeavors not only to enrich the collective knowledge base in NLP but also to lay the groundwork for future investigations and advancements. Ultimately, the aspiration is to pave the way for the development of more robust and dependable LLMs that can navigate the complexities of language generation with a heightened level of accuracy and fidelity.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr).

Puzzle and riddle solving with Large Language Models

Διπλωματική Εργασία 20

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Large Language Models
Τύπος Εργασίας	Survey

Περιγραφή

Puzzle solving and riddle deciphering are cognitive activities deeply intertwined with human intelligence. They serve as intriguing benchmarks for assessing an individual's problem-solving skills and creative thinking, often forming the basis of intelligence quotient (IQ) tests. These tasks demand a multifaceted approach, requiring logical deduction, pattern recognition, lateral thinking, and the ability to draw connections between seemingly disparate pieces of information. The human capacity to excel in these challenges is a testament to the intricacies of our cognitive abilities. However, in the age of Artificial Intelligence (AI) and the proliferation of Large Language Models (LLMs), the landscape of reasoning is undergoing a transformative shift.

The emergence of LLMs has ushered in a new era in the field of reasoning, offering unparalleled potential to tackle intricate puzzles and riddles. While surveys and research papers have delved into the broader domain of reasoning capabilities in LLMs, a conspicuous gap exists when it comes to a dedicated exploration of their application in solving puzzles and riddles. This uncharted territory presents an intriguing opportunity to harness the capabilities of these models in a domain traditionally reserved for human intellect. By undertaking this project, we aim to navigate this field, drawing insights from recent state-of-the-art literature, and shed light on the challenges, opportunities, and limitations that arise when LLMs are employed in the captivating world of puzzles and riddles. Our endeavor promises to contribute to the growing body of knowledge surrounding AI reasoning, providing a comprehensive view of the current landscape and setting the stage for further exploration and refinement of LLMs in this unique cognitive domain.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr).

Robustness in vision-language models

Διπλωματική Εργασία 21

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Explainability
Τύπος Εργασίας	Survey

Περιγραφή

While the recent surge of interest in multimodal pre-training has undeniably yielded valuable insights into the architectural intricacies, available datasets, and training methodologies of multimodal Transformers, the emphasis on robustness and explainability remains an area of critical importance. In an era where these advanced models are being integrated into an ever-expanding array of applications spanning healthcare, autonomous systems, and content recommendation, ensuring their reliability and interpretability is paramount. Robustness, as a foundational pillar of AI development, necessitates a thorough evaluation of how multimodal Transformers perform under diverse and challenging conditions. These conditions could range from noisy data and adversarial attacks to varying contextual cues and domain shifts. Our project seeks to address this knowledge gap by conducting a comprehensive literature review, thereby offering a comprehensive understanding of the robustness landscape within state-of-the-art multimodal Transformers.

Simultaneously, the quest for explainability is an equally compelling imperative in AI research. While multimodal Transformers exhibit exceptional capabilities in processing and understanding multimodal inputs, their decisions must be comprehensible to humans, especially in domains where accountability and trust are paramount. The challenge lies in developing methodologies that can provide insightful explanations for the model's predictions, elucidating the rationale behind its decisions. This literature review project aims to curate and analyze recent techniques that delve into the explainability aspects of multimodal Transformers. By doing so, we aim to contribute to the ongoing dialogue surrounding responsible AI deployment, bridging the gap between sophisticated model performance and the need for transparent and interpretable AI systems, thereby fostering greater trust and acceptance of these transformative technologies across diverse domains and stakeholders.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr).

Knowledge-enhancement of vision-language tasks

Διπλωματική Εργασία 22

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	GNNs, LLMs
Τύπος Εργασίας	Survey

Περιγραφή

The popularity of vision-language tasks, such as visual question answering (VQA) has inspired the extension of the field towards more complex tasks that require knowledge which is not present within the dataset itself, as well as reasoning over such knowledge [1, 2]. For these reasons, datasets such as OK-VQA and A-OKVQA have been developed and demonstrated novel challenges in VL tasks that require external knowledge enhancement. In this thesis, our goal is to improve the performance and the explainability of VL tasks -especially VQA- using external knowledge sources, and specifically both knowledge graphs and large language models. We will evaluate the contribution of each knowledge source in terms of performance, reliability, explainability and extendability through a large amount of experiments.

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, τεχνολογίες και βιβλιοθήκες για transformers. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr) και την Αγγελική Δημητρίου (angelikidim@islab.ntua.gr).

[1] 2303.01903.pdf (arxiv.org)
[2] https://arxiv.org/pdf/2211.12328.pdf

Automatic Generation of Fashion Images Using Prompting in Generative Machine Learning Models

Διπλωματική Εργασία 23

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	LLMs, Generative Models, Reasoning
Τύπος Εργασίας	Research

Περιγραφή

Creative domains like the fashion industry flourish through innovation, self-expression, and a relentless quest for novel aesthetics, which not only captivates individuals and researchers but also intersects with diverse facets of culture, technology, and commerce. Exploring automatic image generation for fashion using machine learning models is intriguing because it enables the creation of diverse and customized fashion content at scale, which aligns with the growing demand for personalized fashion experiences and visual content in the fashion industry. Additionally, leveraging machine learning prompts introduces the potential for fashion-aware AI systems, allowing for contextually relevant and creative image generation based on fashion ontologies with available visual generative models (i.e. Stable Diffusion). The proposed fashion ontology is Fashionpedia [1], a structured representation of fashion concepts and knowledge. Existing literature has addressed the practice of automated prompting, with recent endeavors exploring the integration of ontologies [2]. This thesis presents a substantial challenge, likely demanding significant resources, as it amalgamates multiple contemporary, high-demand themes and techniques, thus necessitating both research and implementation-driven motivation.

[1] https://arxiv.org/pdf/2004.12276.pdf
[2] https://dl.acm.org/doi/pdf/10.1145/3485447.3511921

Απαιτούμενες/επιθυμητές γνώσεις: Python, γνωσεις σε NLP και μοντέλα παραγωγής εικόνων, εξοικείωση με οντολογίες. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr) και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr).

Cluster-aware graph summaries through Graph Matching Neural Networks

Διπλωματική Εργασία 24

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	GNNs
Τύπος Εργασίας	Research

Περιγραφή

Graph summarization is the process of condensing complex graph structures into more manageable and representative forms while retaining crucial information. It allows for efficient analysis of large-scale graphs and aids in uncovering meaningful insights from intricate network data, making it invaluable in various fields. Recent approaches go as far as claiming that the use of representative summaries does not only help in speeding up well-known graph related tasks on heterogeneous graphs (i.e. node classification) but also improves their accuracy [1]. From a XAI point of view, graph summaries offer a promising avenue for identifying meaningful patterns and subgraphs that can be leveraged for insightful explanations.This thesis proposes the creation of graph summaries within specific subgroups, like clusters or labeled classes when available. Such an endeavor would involve a systematic approach to distilling essential information from complex graphs. To this end, exploring graph matching techniques is crucial, enabling the generation of embeddings for entire graphs and individual nodes within the same subgroup. Analyzing these embeddings allows for the identification of the most significant, common, or relevant nodes and edges to form the summaries. This research endeavor is both challenging and research-oriented, requiring a deep dive into Graph Neural Network techniques and experimentation with various approaches to utilize embeddings for subgraph generation. The datasets encompass a range from simpler chemical compounds (MUTAG) to more intricate semantic webs like AIFB, with potential extensions to unlabeled datasets for pattern discovery."

[1] https://dl.acm.org/doi/pdf/10.1145/3487553.3524719

Απαιτούμενες/επιθυμητές γνώσεις: Python, Pytorch, γνωσεις σε θεωρητικό υπόβαθρο και τεχνολογίες για γράφους, νευρωνικά γράφων. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr).

Robustness and Domain Generalisation in Computer Vision using image style manipulation

Διπλωματική Εργασία 25

Επιβλέπων	Θάνος Βουλόδημος
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Robustness, Domain Generalisation
Τύπος Εργασίας	Research

Περιγραφή

This thesis explores the novel application of style transfer techniques in computer vision to enhance the robustness and domain generalisation of deep learning models. It addresses the critical challenge of adapting neural networks to perform reliably across diverse and previously unseen data domains. By incorporating style transfer mechanisms, the research aims to reduce domain-specific biases and enhance the models' ability to recognize and generalise patterns while maintaining their structural integrity. The study investigates the integration of style transfer methods within the training and fine-tuning processes of neural networks, assessing their impact on model resilience in scenarios such as object recognition, image synthesis, and scene understanding. Ultimately, this thesis contributes to the advancement of domain-agnostic computer vision systems capable of accommodating a wide range of visual data, thus fostering their broader applicability in real-world, dynamic environments.

[1] https://arxiv.org/abs/2104.02008
[2] https://arxiv.org/abs/2203.07740

Απαιτούμενες/επιθυμητές γνώσεις: Python, γνωσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: at cs.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Νικόλαο Σπανό (nickspanos23@gmail.com)

Robustness and Domain Generalisation in Computer Vision using adversarial data augmentation

Διπλωματική Εργασία 26

Επιβλέπων	Θάνος Βουλόδημος
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Robustness, Domain Generalisation
Τύπος Εργασίας	Research

Περιγραφή

This thesis delves into the realm of robustness and domain generalization within the field of machine learning by leveraging the power of adversarial data augmentation techniques. The primary focus is to enhance the resilience and adaptability of deep learning models when confronted with variations and shifts in data distributions across different domains. By integrating adversarial data augmentation strategies into the training process, this research aims to create models capable of learning invariant features while effectively reducing the vulnerability to domain-specific biases. The study systematically investigates the application of adversarial networks to generate domain-agnostic synthetic data, which is then seamlessly integrated into the training pipeline to augment the original dataset. Through comprehensive experimentation and evaluation, this thesis seeks to demonstrate the efficacy of this approach in boosting the domain generalization capabilities of machine learning models across various real-world scenarios, thereby contributing to more robust and versatile artificial intelligence systems

[1] https://arxiv.org/abs/2108.02888
[2] https://arxiv.org/abs/2206.07736

Απαιτούμενες/επιθυμητές γνώσεις: Python, γνωσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: at cs.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Νικόλαο Σπανό (nickspanos23@gmail.com)

Robustness and Domain Generalisation in Computer Vision using contrastive learning

Διπλωματική Εργασία 27

Επιβλέπων	Θάνος Βουλόδημος
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Robustness, Domain Generalisation
Τύπος Εργασίας	Research

Περιγραφή

The thesis explores the powerful synergy of contrastive learning and pretraining techniques to enhance the robustness and domain generalisation capabilities of deep learning models. It addresses the persistent challenge of enabling neural networks to perform effectively across diverse and previously unseen data domains. By leveraging the power of contrastive learning, the research aims to create representations that capture underlying semantic structures and reduce the domain gap. The study systematically investigates the incorporation of pretraining strategies, such as transfer learning and self-supervised learning, in conjunction with contrastive learning to initialize and fine-tune models. It assesses their impact on model adaptation and generalisation across various domains, including natural language processing, computer vision, and audio analysis. Ultimately, this thesis contributes to the development of domain-agnostic machine learning systems capable of efficiently tackling a wide array of real-world challenges, thereby advancing the field's capacity to deploy robust and versatile artificial intelligence solutions.

[1] https://arxiv.org/abs/2103.16050
[2] https://arxiv.org/abs/2303.01289

Απαιτούμενες/επιθυμητές γνώσεις: Python, γνωσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: at cs.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Νικόλαο Σπανό (nickspanos23@gmail.com)

Utilizing Diffusion Models for generation of tabular data

Διπλωματική Εργασία 28

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Diffusion Models, Generative Models
Τύπος Εργασίας	Research

Περιγραφή

Diffusion models have emerged as the new state of the art family of deep generative models. They have broken the long-standing dominance of Generative Adversarial Networks (GANs) in the challenging task of image generation.

These models are a family of probabilistic generative models that progressively destroy the training data by diffusing/introducing noise. They then learn to reverse this process, in order to generate new samples.

Despite the widespread application of diffusion models in computer vision, one area that remains largely unexplored is their extension to the field of tabular data. Tabular data remains the most prevalent data type in both research and industry. The production of high quality synthetic tabular data presents difficulties due to the heterogeneity found in their characteristics (continuous and discrete numerical data, categorical data) but is of significant importance in an era governed by strict privacy regulations. It is therefore becoming increasingly critical as it allows researchers and organizations to utilize and share valuable information without compromising individual privacy.

This thesis aims to delve into the uncharted territory of applying diffusion models to tabular data problems and investigate their performance on different datasets.

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch). Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Λευτέρη Τσώνη (e-mail: lefteristsonis at outlook.com)

[1] Yang, L. and Zhang, Z. et al. (2022) 'Diffusion Models: A Comprehensive Survey of Methods and Applications'

[2] Kotelnikov, A. et al. (2022) 'TabDDPM: Modelling Tabular Data with Diffusion Models'

Medical Image applications of Diffusion Models

Διπλωματική Εργασία 29

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Diffusion Models, Generative Models, Medical Imaging
Τύπος Εργασίας	Research

Περιγραφή

Diffusion Models, a family of deep generative models, have gained widespread popularity for their ability to model complex distributions and generate data. They excel at generating synthetic data samples, often setting the state of the art in computer vision, among other tasks. However, their untapped potential in medical image analysis could bring significant advancements to healthcare.

These models have risen to prominence due to their capacity to capture complex data patterns. In computer vision, for instance, they excel at enhancing image quality and synthesizing new samples. In medical imaging, where precision is crucial, diffusion models could revolutionize diagnosis and treatment planning by improving image quality, reducing noise, and aiding in detecting subtle anomalies. They can also generate synthetic images to train deep learning algorithms when labeled data is limited, a common phenomenon in the medical domain, improving overall system performance.

In this thesis, we will explore the various applications of diffusion models in medicine. Exploring their use in medical image analysis has the potential to enhance healthcare by improving image quality, aiding diagnosis, and advancing our understanding of complex diseases, benefiting both patients and healthcare professionals.

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch). Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Λευτέρη Τσώνη (e-mail: lefteristsonis at outlook.com)

[1] Yang, L. and Zhang, Z. et al. (2022) ‘Diffusion Models: A Comprehensive Survey of Methods and Applications’

[2] Pinaya W.H.L. et al. (2023) ‘Generative AI for Medical Imaging: extending the MONAI Framework’

Incorporating sketch guidance in Text-to-Image Diffusion Models

Διπλωματική Εργασία 30

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Diffusion Models, Generative Models
Τύπος Εργασίας	Research

Περιγραφή

Diffusion models have emerged as the new state of the art family of deep generative models. They have broken the long-standing dominance of Generative Adversarial Networks (GANs) in the challenging task of image generation.

These models are a family of probabilistic generative models that progressively destroy the training data by diffusing/introducing noise. They then learn to reverse this process, in order to generate new samples.

Although diffusion models now outperform older methods in image synthesis benchmarks, their application to the Sketch-to-Image Translation task has not been thoroughly explored. This task involves the synthesis of realistic images, driven by text prompts, while additionally being subject to the constraints imposed by a sketch provided as input to the model, during inference time.

In this thesis, we will begin by studying the theory behind diffusion models and then we will perform experimental tests and comparisons, exploiting diffusion models in the Sketch-to-Image Translation task.

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch). Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Λευτέρη Τσώνη (e-mail: lefteristsonis at outlook.com)

[1] Yang, L. and Zhang, Z. et al. (2022) ‘Diffusion Models: A Comprehensive Survey of Methods and Applications’

[2] Voynov, A. et al. (2022) ‘Sketch-Guided Text-to-Image Diffusion Models’

Use of ViT (Vision Transformers) in low-data classification applications with modern data augmentation methods

Διπλωματική Εργασία 31

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Transformers,Robustness
Τύπος Εργασίας	Research

Περιγραφή

Transformers and their specialized variant, Vision Transformers (ViT), have revolutionized the field of artificial intelligence, especially in the areas of natural language processing and computer vision. What makes Transformers innovative is their attention mechanism, which allows them to process sequences of data in parallel, capturing complex dependencies and relationships between elements. In the case of ViT-Transformers, the innovation was extended to the field of computer vision. Rather than relying on traditional convolutional neural networks (CNNs), ViTs harness the power of self-awareness to process images as sequences of patches, allowing them to outperform CNNs in various visual recognition tasks. The innovation of Transformers and ViT lies not only in their impressive performance, but also in their flexibility, as they have adapted to a wide range of applications, expanding the limits of machine learning and artificial intelligence capabilities. The aim of this thesis is to use modern data augmentation methods to improve the performance of ViT-Transformers in applications where the data volume is insufficient and to investigate how the front-end mechanism responds to complex transformations of input images.

[1]How to train your ViT? Data, Augmentation, and Regularization in Vision Transformers, Transactions on Machine Learning Research (https://arxiv.org/abs/2106.10270)
[2]AutoMix: Unveiling the Power of Mixup for Stronger Classifiers,ECCV 2022 (https://arxiv.org/abs/2103.13027)
[3]PixMix: Dreamlike Pictures Comprehensively Improve Safety Measures, CVPR 2022 (https://arxiv.org/abs/2112.05135)

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch). Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr), Νικόλαο Σπανό (nickspanos23@gmail.com)

Use of ViT (Vision Transformers) in low-data segmentation applications with modern data augmentation methods

Διπλωματική Εργασία 32

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Transformers,Robustness
Τύπος Εργασίας	Research

Περιγραφή

In segmentation tasks, Vision Transformers (ViT) have emerged as a transformative technology, offering a new perspective to image understanding.By adapting the self-attention mechanism to handle pixel-level predictions, ViT models excel in semantic and instance segmentation tasks.They exhibit the ability to capture both local and global context information, enabling accurate delineation of object boundaries and precise pixel-level labeling. This innovation has paved the way for more efficient and accurate image segmentation, significantly benefiting applications ranging from medical imaging to autonomous driving and beyond. However, in many applications the volume of data is not sufficient to enable training to be efficient in transformer architectures.The purpose of this thesis is to investigate complex augmentation methods data in combination with attention mechanisms to improve the generalization of such models in image segmentation applications.

[1] How to train your ViT? Data, Augmentation, and Regularization in Vision Transformers, Transactions on Machine Learning Research (https://arxiv.org/abs/2106.10270)
[2] SegViT: Semantic Segmentation with Plain Vision Transformers, NeurIPS 2022 (https://arxiv.org/abs/2210.05844)
[3] MedViT: A Robust Vision Transformer for Generalized Medical Image Classification (https://arxiv.org/abs/2302.09462)
[4] PixMix: Dreamlike Pictures Comprehensively Improve Safety Measures, CVPR 2022 (https://arxiv.org/abs/2112.05135)

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch). Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr), Νικόλαο Σπανό (nickspanos23@gmail.com)

Using Contrastive Learning to Detect AI-Generated Data

Διπλωματική Εργασία 33

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Contrastive Learning
Τύπος Εργασίας	Research

Περιγραφή

Contrastive learning is a cutting-edge approach to machine learning that has garnered significant attention due to its innovation and effectiveness. Unlike traditional supervised learning, where data is labeled, contrastive learning does not require labeled data for training. Instead, it leverages unlabeled data and aims to learn meaningful representations through contrasting positive pairs (similar examples)and negative pairs (dissimilar examples). This self-supervised learning technique has found promising applications in various domains,such as computer vision and natural language processing.One notable application is the detection of material generated by artificial intelligence,such as deepfake videos and synthetic texts. By extracting complex features from these generated contents and comparing them with authentic data, counterfactual learning can play a critical role in identifying and mitigating the dissemination of misleading or malicious material that produced by AI, thus addressing a pressing problem in the era of digital disinformation.The aim of this thesis is to apply modern methods of detecting such data and their further improvement.

[1] ConDA: Contrastive Domain Adaptation for AI-generated Text Detection (https://paperswithcode.com/paper/conda-contrastive-domain-adaptation-for-ai)
[2] GenImage: A Million-Scale Benchmark for Detecting AI-Generated Image (https://arxiv.org/abs/2306.08571)
[3] On The Detection of Synthetic Images Generated by Diffusion Models (https://ieeexplore.ieee.org/abstract/document/10095167?casa_token=t5hjZ5PCaLwAAAAA:cB8CNxJy6ND4ldeWoZ-KC94RLYdk40u_PVeATNqAgAaMz2s66a8VGuQdWlrMa33mBHLbdUjPJS3b)

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch). Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr), Νικόλαο Σπανό (nickspanos23@gmail.com)

Interpretable Vision Transformers

Διπλωματική Εργασία 34

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Computer Vision, Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

Vision transformers (ViTs), which have demonstrated a state-of-the-art performance in image classification, can also visualize global interpretations through attention-based contributions. However, the complexity of the model makes it difficult to interpret the decision-making process, and the ambiguity of the attention maps can cause incorrect correlations between image patches. For this reason, new systems are proposed to not only improve the classification performance but also provide a human-friendly interpretation, which is effective in resolving the trade-off between performance and interpretability. The purpose of this task is the study of architectures and the investigation of methods that can add interpretability into ViT-based systems, improving their performance.

[1] https://arxiv.org/pdf/2010.11929.pdf

[2] https://proceedings.mlr.press/v162/kim22g/kim22g.pdf

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch, TensorFlow). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv@mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr)

Medical Image Segmentation

Διπλωματική Εργασία 35

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Computer Vision, Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

Medical image segmentation using Vision Transformers (ViTs) represents a cutting-edge approach in the field of healthcare and diagnostic imaging. ViTs, originally designed for natural image analysis, have demonstrated their adaptability to the unique challenges of medical imaging. By leveraging their ability to capture intricate spatial patterns and contextual information, ViTs excel at delineating anatomical structures, lesions, and pathologies in medical images, such as MRI and CT scans. By segmenting these images into distinct regions or objects of interest, medical professionals can gain invaluable insights for diagnosis, treatment planning, and monitoring the progression of diseases. Medical image segmentation has a wide range of applications, from tumor detection and organ localization to vascular analysis and neuroimaging. As this field continues to advance, ViTs hold promise in revolutionizing the way healthcare professionals interpret and utilize complex medical imagery, ultimately improving patient care and outcomes. The purpose of this study is the investigation of methods used for image segmentation, especially in tasks referred to medical imaging.

[1] https://www.sciencedirect.com/science/article/pii/S1746809423002240

[2] https://www.sciencedirect.com/science/article/pii/S1361841523000634

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch, TensorFlow). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv@mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr)

Semantic Segmentation for Autonomous Driving

Διπλωματική Εργασία 36

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Computer Vision, Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

Semantic segmentation using Vision Transformers (ViTs) in the context of autonomous driving represents a cutting-edge approach to scene understanding and perception. By applying ViTs to high-resolution camera images, autonomous vehicles can achieve fine-grained semantic segmentation, enabling the precise identification and categorization of objects and road elements in the environment. This technology empowers self-driving cars to distinguish between vehicles, pedestrians, road signs, lanes, and other critical elements on the road. With its capacity for capturing intricate spatial relationships, ViTs can enhance the vehicle's ability to navigate safely, plan optimal routes, and make real-time decisions, thus contributing to the advancement of autonomous driving systems' accuracy and safety. The purpose of this study is the investigation of methods used for semantic segmentation, especially in tasks referred to autonomous driving.

[1] https://www.sciencedirect.com/science/article/pii/S0925231222000054

[2] https://www.sciencedirect.com/science/article/pii/S0952197623008539

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch, TensorFlow). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv@mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr)

Small Object Detection in UAV-Vision

Διπλωματική Εργασία 37

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Computer Vision, Explainable AI
Τύπος Εργασίας	Research

Περιγραφή

Small object detection in Unmanned Aerial Vehicle (UAV) vision is a critical and complex task with a wide range of applications, including search and rescue, surveillance, agriculture, and environmental monitoring. UAVs equipped with high-resolution cameras can capture detailed images from various altitudes. However, detecting small objects, such as individual persons, animals, or specific objects of interest, in these images poses unique challenges due to their size, scale variations, and potential occlusions. Advanced computer vision algorithms and deep learning models are employed to address these challenges, allowing UAVs to autonomously identify and track small objects with precision. The successful detection of small objects in UAV vision not only enhances the capabilities of autonomous systems but also plays a vital role in improving safety, security, and efficiency across numerous domains. The purpose of this study is the investigation of methods used for object detection centered on small and tiny objects derived from UAV-Vision tasks.

[1] https://www.sciencedirect.com/science/article/pii/S1051200422001312

[2] https://www.sciencedirect.com/science/article/pii/S0262885622001007

[3] https://ieeexplore.ieee.org/abstract/document/10168277

[4] https://ieeexplore.ieee.org/abstract/document/10035490

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch, TensorFlow). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv@mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr)

Advanced methods for Image Captioning

Διπλωματική Εργασία 38

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Advanced methods for image captioning leverage cutting-edge techniques in deep learning, multimodal modeling, and natural language processing to generate rich and contextually relevant textual descriptions for images. These methods often combine the power of pretrained multimodal models with attention mechanisms and transformer architectures to capture intricate relationships between visual content and language. They excel in understanding complex scenes, handling multiple objects, and generating coherent and coherent captions. Additionally, advanced image captioning models are capable of zero-shot and few-shot learning, enabling them to describe images without task-specific training data. These innovations have made advanced image captioning methods invaluable in applications ranging from accessibility for the visually impaired to content indexing, multimedia presentations, and beyond, transforming the way we interact with and understand visual content. The purpose of this study is the investigation of new methods used for image captioning and their application in different tasks.

[1] https://link.springer.com/article/10.1007/s10462-023-10488-2

[2] https://www.ieee-jas.net/en/article/id/dbb31c0b-399b-4d21-a8ec-aa33ff6f87af

[3]https://www.researchgate.net/publication/369056656_A_Survey_on_Attention-Based_Models_for_Image_Captioning

[4] https://arxiv.org/abs/2205.01917

[5] https://arxiv.org/abs/2205.14100

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch, TensorFlow). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv@mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr)

Spatio-temporal models for Skeleton-based Action Recognition

Διπλωματική Εργασία 39

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Computer Vision
Τύπος Εργασίας	Research

Περιγραφή

Spatio-temporal models play a pivotal role in skeleton-based action recognition tasks, where they excel at capturing the intricate interplay of spatial and temporal cues within human pose data. These models are designed to process sequences of skeletal joint positions over time, effectively encoding the dynamics of human actions. Leveraging both the spatial relationships between skeletal joints and the temporal dynamics of these joints across consecutive frames in video sequences, these models offer a holistic understanding of human movement. Through techniques like Graph Convolutional Networks (GCNs), Recurrent Neural Networks (RNNs) and Transformers, spatio-temporal models extract and propagate spatial information between joints while modeling the temporal evolution of these relationships frame by frame. This enables them to recognize a wide range of complex actions, from basic gestures to intricate athletic movements, with high accuracy. Spatio-temporal models have significantly advanced the field of action recognition, finding applications in diverse domains like healthcare, robotics, and entertainment, where understanding and interpreting human actions are essential for intelligent decision-making and interaction.

[1]https://link.springer.com/article/10.1007/s11263-022-01594-9

[2]https://ieeexplore.ieee.org/abstract/document/9795869?casa_token=PtvcGLYtGBUAAAAA:gIuIctKxGBRTYAh0RVpaf5lz4LoJCuwaTOrDlaH-Clnb8SELAhtDI0kNaYsZVIhPq0X0NjSFLw

[3] https://paperswithcode.com/task/skeleton-based-action-recognition

[4]https://openaccess.thecvf.com/content/CVPR2022/papers/Duan_Revisiting_Skeleton-Based_Action_Recognition_CVPR_2022_paper.pdf

[5] https://arxiv.org/pdf/2012.06399.pdf

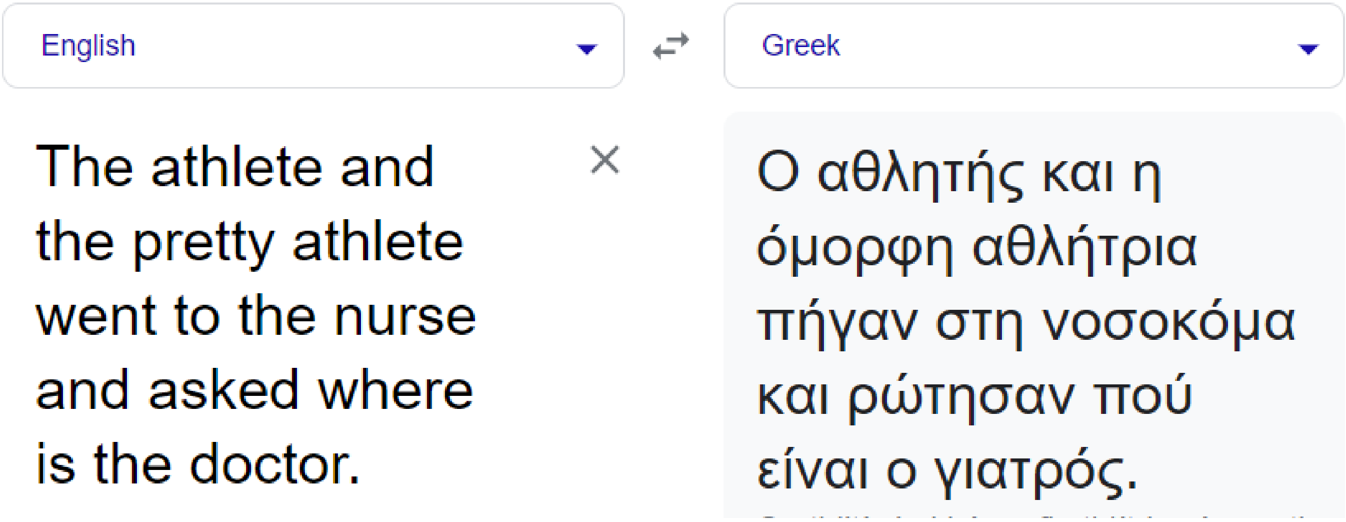
Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch, TensorFlow). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv@mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr)

Detecting and Evaluating Gender Bias in Machine Translation

Διπλωματική Εργασία 40

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
	Διαθέσιμη
Status	NLP/Biases
Περιοχή	Research
Τύπος Εργασίας	
Επιβλέπων	Γιώργος Στάμου

Περιγραφή



The athlete and the pretty athlete went to the nurse and asked where is the doctor.

Ο αθλητής και η όμορφη αθλήτρια πήγαν στη νοσοκόμα και ρώτησαν πού είναι ο γιατρός.

Machine Translation (MT) systems, being trained on large corpora of human-generated text, can inadvertently learn and perpetuate societal biases present in these texts. Gender bias in MT is a particularly concerning issue, given that language plays a pivotal role in shaping societal perceptions and norms. For instance, when translating a gender-neutral sentence from a language like English into Greek or Spanish, the MT system might introduce unwarranted gender specifications, leading to biased translations. Understanding, detecting, and evaluating such biases is not just crucial for improving translation accuracy, but also for fostering equity and fairness in language technologies.

Description of Work:

1. **Literature Review:** Conduct an extensive review of current literature on biases in machine learning, focusing specifically on gender biases in MT systems. This will include understanding existing metrics, evaluation techniques, and debiasing methodologies in the domain.

2. **Bias Detection:** Develop a methodology to automatically detect potential gender biases in translations. This could involve:
    a. Analyzing cases where gender is introduced in the target language when it was absent in the source.
    b. Identifying scenarios where gender is changed or inferred differently in the translation process.
3. **Evaluation Metric Creation:** Design a robust evaluation metric that quantifies gender bias in translations. This would provide a standardized way to assess different MT systems in terms of their gender bias.
4. **Model Analysis:** Using the developed detection methodology and evaluation metric, analyze popular MT systems to gauge their level of gender bias. This will provide insights into how prevalent such biases are in state-of-the-art systems.
5. *[Optional]* **Bias Mitigation Recommendations:** Based on the findings, propose potential mitigation strategies to reduce gender bias in MT systems. This could involve techniques like retraining models with debiased data, post-processing translations, or integrating gender-neutral alternatives into translation outputs.
6. **Validation and Testing:** Validate the effectiveness of the bias detection methodology and the evaluation metric on held-out datasets, ensuring their robustness and generalizability across diverse texts and languages.

Datasets and Tasks:

Parallel Corpora: Utilize parallel corpora, like the ones from the WMT (Workshop on Machine Translation) challenges, which contain sentences in source languages and their corresponding translations. Analyzing these can help identify instances of introduced or altered gender.

Related Literature:

[1] Savoldi, Beatrice, et al. "Gender bias in machine translation." Transactions of the Association for Computational Linguistics 9 (2021): 845-874.

[2] Bolukbasi, Tolga, et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." Advances in neural information processing systems 29 (2016).

[3] Stanovsky, Gabriel, Noah A. Smith, and Luke Zettlemoyer. "Evaluating gender bias in machine translation." arXiv preprint arXiv:1906.00591 (2019).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες μηχανικής μάθησης, εξοικείωση με τεχνολογίες εκμάθησης νευρωνικών δικτύων και με συστήματα επεξεργασίας φυσικής γλώσσας (NLP). Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τον Ορφέα Μενή - Μαστρομιχαλάκη (e-mail: menorf at ails.ece.ntua.gr) και τον Γιώργο Φιλανδριανό (email: geofila at islab.ntua.gr).

Visual Counterfactuals on Medical Images

Διπλωματική Εργασία 41

Επιβλέπων	Γιώργος Στάμου
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Vision, Explainable AI
Τύπος Εργασίας	Reproduction, Research

Περιγραφή

In the last few years, the medical community is showing increasing interest in utilizing Machine Intelligence in the process of patient screening and diagnosis. However, such a critical field that deals with human lives requires a clear understanding of how any AI system utilized makes decisions. Systems based on Symbolic AI have been used successfully since the 60's in the medical field, using patient symptoms to make diagnoses. Such systems though are not capable of dealing with high-dimensional data such as images, while Deep Learning methods are showing outstanding success.

Despite their success, one significant drawback of Deep Learning methods, in contrast with Symbolic AI systems, is that they are not inherently interpretable. This raises issues of trust by both doctors who use these systems as aides and patients who receive such diagnoses, and results in (justifiably) limited adoption. The goal of the eXplainable AI (XAI) field is to provide some understanding on the inner workings of such systems in order to build trust and avoid biased decisions. A technique with growing popularity is Counterfactual Explanations (CE) which provide explainability by highlighting which parts of the input needed to be different in order for the output to be different. In the context of medical images CEs would answer questions such as “Why was this tumor classified as benign?”, “Which visual characteristics would make the AI system to classify it as malignant?”. Many techniques of CEs on tabular data have been developed and adopted successfully on relevant data, but techniques operating on images are still lacking and usually applied to toy-datasets such as MNIST.

The goal of this project is to reproduce and apply existing techniques on demanding medical datasets and evaluate their effectiveness. Optionally, the insights gained by the reproduction can then be used to build new techniques addressing the weaknesses of existing methods.

Description of work:

1. Brief survey of existing CE methods for images.
2. Reproduction of one or more of these methods on the datasets used in the original papers.
3. Reproduction of the same methods on real-world medical images.
4. Assessment of the robustness of these methods by contrasting their behavior on simple and complex images.

5. Development of new techniques based on the insights gained.

[1] Guidotti, Riccardo. "Counterfactual explanations and how to find them: literature review and benchmarking." *Data Mining and Knowledge Discovery* (2022): 1-55.
[2] Zhao, Yunxia. "Fast real-time counterfactual explanations." *arXiv preprint arXiv:2007.05684* (2020).
[3] Vermeire, Tom, et al. "Explainable image classification with evidence counterfactual." Pattern Analysis and Applications 25.2 (2022): 315-335.

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες Deep Learning (π.χ. PyTorch). Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam@cs.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr), τον Ιάσονα Λιάρτη (e-mail: jliartis@ails.cs.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr).

Neuro-Symbolic AI for Image Classification

Διπλωματική Εργασία 42

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Vision, Neuro-Symbolic AI, Explainable AI
Τύπος Εργασίας	Survey, Research

Περιγραφή

In the field of Computer Vision, a typical Deep Neural Network (DNN) will consist of two main blocks, the feature extractor, usually in the form of a Convolutional Neural Network or a Vision Transformer, and the classifier, which is typically a Multi-Layer Perceptron (MLP). DNNs have been very successful due to their outstanding ability to learn good, high-level visual features from data without the need for human guidance. The caveat is that these features, in contrast to the hand-crafted features of classical Computer Vision, are not easily mapped to human-understandable concepts, and furthermore, the way these features are used by the MLP is completely opaque. The field of Neuro-Symbolic Artificial Intelligence (NeSyAI) attempts to fuse the success of feature learning provided by DNNs with the inherent interpretability of Symbolic Expressions, stated in human-understandable concepts and in natural language.

Different methods have been proposed to train DNNs to inherently learn human-understandable representations or to assign such human-understandable concepts to the features learned after the training has concluded. These concepts are then passed to some symbolic module that extracts symbolic expressions used to classify these images. Many different approaches have been used for the symbolic modules, differing in the language of the symbolic expressions (such as propositional logic, first-order logic, and answer set programs) and the way the expressions are reverse-engineered.

The goal of this project is to conduct a survey of recent methods and to contrast them based on the components they deploy, the way they are trained and the datasets they have been successfully applied to. Optionally, original methods can be developed which utilized the under-explored use of Ontologies, Description Logic concepts and Conjunctive Queries.

Description of work:

1. Collection of existing surveys on this topic.
2. Collection of recent published work.
3. Organized summarization of these works in the form of a literature survey.
4. Development of novel methods utilizing some combination of Ontologies, Description Logic concepts and Conjunctive Queries.

[1] Stammer, Wolfgang, Patrick Schramowski, and Kristian Kersting. "Right for the right concept: Revising neuro-symbolic concepts by interacting with their explanations." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.
[2] Parth Padalkar, et al. “NeSyFOLD: Neurosymbolic Framework for Interpretable Image Classification.” arXiv preprint arXiv:2301.12667 (2023).
[3] Shindo, Hikaru, et al. "α ILP: thinking visual scenes as differentiable logic programs." *Machine Learning* 112.5 (2023): 1465-1497.
[4] Marconato, Emanuele, et al. "Neuro symbolic continual learning: Knowledge, reasoning shortcuts and concept rehearsal." *arXiv preprint arXiv:2302.01242* (2023).

Απαιτούμενες/επιθυμητές γνώσεις: Βαθιά Νευρωνικά Δίκτυα, Τεχνητή Νοημοσύνη, εξοικείωση με κάποιας μορφής λογική. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam@cs.ntua.gr), τον Ιάσονα Λιάρτη (e-mail: jliartis@ails.cs.ntua.gr), και τον Ορφέα Μενή - Μαστρομιχαλάκη (menorf at ails.ece.ntua.gr).

Semantic Enrichment of Categorical Features

Διπλωματική Εργασία 43

Επιβλέπων	Γιώργος Στάμου
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Machine Learning, Ontological Knowledge Representation, Natural Language Processing, Explainable Artificial Intelligence
Τύπος Εργασίας	Survey, Research

Περιγραφή

A very large percentage of real-word data is tabular, and within this tabular data a lot of features are categorical, i.e. they are drawn from a limited collection of values. Most popular Machine Learning models (such as Neural Networks and Support Vector Machines) are not inherently compatible with categorical features since they only accept numeric values as input. The usual and crude solution is to transform categorical features to numeric by a simple one-hot encoding. Even when a model can accept categorical features (such as some Decision Tree algorithms) the values of a categorical feature are treated as completely distinct from one another. Take this credit score dataset as an example. One column lists the occupation of the person. Is a doctor as dissimilar to a scientist as they are to an entrepreneur? Shouldn't an algorithm that deals with these values consider some of them more similar than others?

When treating these values in the aforementioned manners, the distance between two values is zero when the values are identical or some constant if they are different. In this modus operandi all the meaning we associate with different words such as "doctor" and "scientist" is lost. To combat this shortcoming a plethora of alternatives have been studied which enrich different categorical values with the real-world meaning these different values possess.

The goal of this project is to conduct a survey on existing methods of semantically enriching categorical data. These methods usually either employ technologies from Ontologies, which have word hierarchies and relations curated by humans (such as WordNet) or learned word embeddings from Language Models. These different methods will be contrasted, including their usage across different tasks such as calculating distances for clustering or Nearest Neighbor classification. Optionally, they will be employed in novel ways in the field of Explainable Artificial Intelligence where often instances must be evaluated on their similarity in order to construct explanations (see Prototypes and Criticisms, and Counterfactual Explanations).

Description of work:

1. Collection of existing surveys on this topic.
2. Collection of recent published work.

3. Organized summarization of these works in the form of a literature survey.
4. Novel application of these methods in the field of Explainable Artificial Intelligence.

[1] Sarasvananda, Ida Bagus Gede, Retantyo Wardoyo, and Anny Kartika Sari. "The k-means clustering algorithm with semantic similarity to estimate the cost of hospitalization." IJCCS (Indonesian Journal of Computing and Cybernetics Systems) 13.4 (2019): 313-322.
[2] Cerda, Patricio, Gaël Varoquaux, and Balázs Kégl. "Similarity encoding for learning with dirty categorical variables." Machine Learning 107.8-10 (2018): 1477-1494.
[3] Mumtaz, Summaya, and Martin Giese. "Frequency-Based vs. Knowledge-Based Similarity Measures for Categorical Data." AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering (1). 2020. (pdf)

Απαιτούμενες/επιθυμητές γνώσεις: Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Βαθιά Μάθηση. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam@cs.ntua.gr), και τον Ιάσονα Λιάρτη (e-mail: jliartis@ails.cs.ntua.gr).

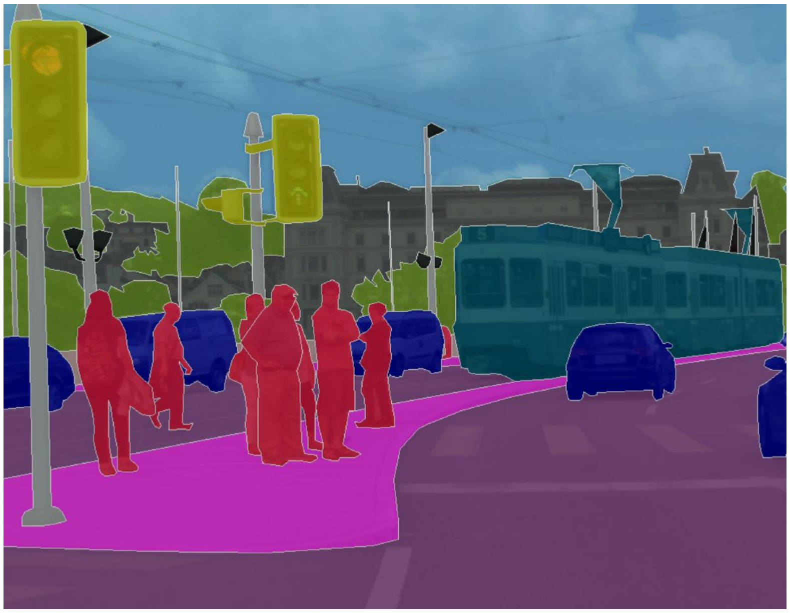
Fast Data Annotation Utilizing Automatic Image Segmentation

Διπλωματική Εργασία 44

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Vision, Software Engineering, Knowledge Engineering
Τύπος Εργασίας	Dataset Creation

Περιγραφή

In the field of Computer Vision, one often encounters problems where a single label characterizing the whole image is not enough. One such crucial task is object segmentation and detection which is needed to train autonomous driving systems to detect cars, pedestrians, traffic lights, stop signs etc. and react accordingly. To train such systems, one needs images where each object is annotated with an outline and a label categorizing it as a car, a pedestrian etc. Such richly annotated images also find many applications in the fields of Explainable Artificial Intelligence and Neuro-Symbolic Artificial Intelligence where pixel level information is not enough.



Yet, public, openly accessible datasets containing such images are scarce. The most significant reason why is that creating these datasets is very laborious since the process cannot easily be automated. Usually humans need to inspect each image, draw the outline of each object by hand and assign them labels. Even after this process is concluded more humans are needed to assess whether the annotations that have been produced are of sufficient quality and do not contain mistakes.

This project aims to utilize recent developments in image segmentation technologies in order to automate the process of outlining each object without sacrificing quality. After a robust pipeline has been produced, crowdsourcing will be used in order to create a sizable dataset. A very important part of this work will be developing a good UI that allows users to annotate objects easily and efficiently and rate that automatically produced outlines in order to maintain a good quality for the dataset.

Description of work:

1. Collection of recent works on image segmentation.
2. Deployment of a selected model considering hardware requirements and accuracy.
3. Development of a website with a quality UI that allows users to annotate objects.
4. Creation of a crowdsourcing campaign to collect user annotations and ratings.
5. Publication of the collected dataset on an open repository such as Kaggle.

[1] https://www.cityscapes-dataset.com
[2] https://cocodataset.org
[3] https://www.kaggle.com/datasets
[4] Kirillov, Alexander, et al. "Segment anything." arXiv preprint arXiv:2304.02643 (2023).

Απαιτούμενες/επιθυμητές γνώσεις: Python, βιβλιοθήκες Deep Learning (π.χ. PyTorch), web development, software engineering. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam@cs.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar@image.ece.ntua.gr), τον Ιάσονα Λιάρτη (e-mail: jliartis@ails.cs.ntua.gr) και την Παρασκευή Θεοφίλου (e-mail: paristh@ails.ece.ntua.gr).

Learning to Reason from Data

Εργασία Μαθήματος 1

Επιβλέπων	Γιώργος Στάμου
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Automated Reasoning, Machine Learning, Natural Language Processing
Τύπος Εργασίας	Reproduction

Περιγραφή

This project is a reproduction of the paper “On the Paradox of Learning to Reason from Data”.

Abstract:

Logical reasoning is needed in a wide range of NLP tasks. Can a BERT model be trained end-to-end to solve logical reasoning problems presented in natural language? We attempt to answer this question in a confined problem space where there exists a set of parameters that perfectly simulates logical reasoning. We make observations that seem to contradict each other: BERT attains near-perfect accuracy on in-distribution test examples while failing to generalize to other data distributions over the exact same problem space. Our study provides an explanation for this paradox: instead of learning to emulate the correct reasoning function, BERT has in fact learned statistical features that inherently exist in logical reasoning problems. We also show that it is infeasible to jointly remove statistical features from data, illustrating the difficulty of learning to reason in general. Our result naturally extends to other neural models and unveils the fundamental difference between learning to reason and learning to achieve high performance on NLP benchmarks using statistical features.

[1] Zhang, Honghua, et al. "On the Paradox of Learning to Reason from Data, May 2022." URL http://arxiv.org/abs/2205.11502.

Απαιτούμενες/επιθυμητές γνώσεις: Python, PyTorch, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Βαθιά Μάθηση. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam@cs.ntua.gr), και τον Ιάσονα Λιάρτη (e-mail: jliartis@ails.cs.ntua.gr).

Image dataset from time-series EEG signals

Εργασία Μαθήματος 2

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Computer Vision
Τύπος Εργασίας	Dataset Creation

Περιγραφή

In EEG analysis, due to the nature of the signals and the inherent difficulty in identifying anomalies from them, it is common to convert them into images in order to encode the information into more accessible form. This step is commonly performed as a pre-processing step and, as a result, there are no ready-made data sets that an engineer can process. The objective of this work is to create certain datasets from time series of encephalograms using known methods in research, e.g. Mel-Spectrograms, Gramian Angular Fields, Recurrence plots and to research by training deep learning models on which transformations offer the best performance for certain applications, such as epilepsy detection in encephalograms.

[1] Implementation of Deep Neural Networks to Classify EEG Signals using Gramian Angular Summation Field for Epilepsy Diagnosis (https://arxiv.org/abs/2003.04534)
[2] A list of all public EEG-datasets (https://github.com/meagmohit/EEG-Datasets)

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) , Νικόλαο Σπανό (nickspanos23@gmail.com)

Cracking the Code: Hands-On use of Diffusion Models for Image Generation

Εργασία Μαθήματος 3

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Diffusion Models, Generative Models
Τύπος Εργασίας	Reproduction

Περιγραφή

Diffusion models have emerged as the state of the art category of deep generative models, toppling the longstanding dominance of Generative Adversarial Networks (GANs) in the challenging task of image generation. Essentially, they are probabilistic generative models that initiate the training process by progressively incorporating noise into the training dataset. As this iterative procedure unfolds, the models acquire the ability to reverse the diffusion process, enabling them to produce entirely novel samples starting from pure noise.

In the context of this assignment, after getting to know the basics of diffusion model theory, we will perform hands-on experiments, using pre-trained diffusion models to generate synthetic images. We will use various datasets to evaluate and compare the performance of these models using different benchmarks. Additionally, we will explore how various hyperparameters affect sample quality.

Through these experiments, you will form a better understanding of the potential and the limitations of diffusion models. Furthermore, you will hone your ability to reproduce code and utilize popular machine learning frameworks.

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Λευτέρη Τσώνη (e-mail: lefteristsonis at outlook.com)

[1] Yang, L. and Zhang, Z. et al. (2022) ‘Diffusion Models: A Comprehensive Survey of Methods and Applications’

[2] Karras, T. et al. (2022) ‘Elucidating the Design Space of Diffusion-Based Generative Models’

Rising to the summit: Diffusion Models as state of the art deep generative models

Εργασία Μαθήματος 4

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Diffusion Models, Generative Models
Τύπος Εργασίας	Survey

Περιγραφή

This assignment delves into the fundamental theory underpinning diffusion models, an emerging category of probabilistic generative models that have recently garnered significant attention. Our primary objective is to conduct a thorough exploration of the theoretical foundations of diffusion models and craft a concise survey encapsulating the current body of research within this domain.

We will begin with a meticulous literature review to pinpoint pivotal research contributions in diffusion models. This class of deep generative models has witnessed remarkable advancements in the past three years, making it all the more intriguing to chronologically trace the key milestones in its development.

Our ultimate goal is to produce a meticulously organized survey that captures the essence of diffusion model theory, as well as the many applications these models find. This survey promises to be a valuable resource, offering a foundational understanding of these models to anyone venturing into this field.

Απαιτούμενες/επιθυμητές γνώσεις: Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Λευτέρη Τσώνη (e-mail: lefteristsonis at outlook.com)

[1] Yang, L. and Zhang, Z. et al. (2022) ‘Diffusion Models: A Comprehensive Survey of Methods and Applications’

Creative Chaos: Prompting Diffusion Models to their limits

Εργασία Μαθήματος 5

Επιβλέπων	Θάνος Βουλόδημος
Συνεπιβλέπουσα	Παρασκευή Τζούβελη
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Diffusion Models, Generative Models
Τύπος Εργασίας	Prompting

Περιγραφή

Diffusion models have risen to prominence as the state of the art family of deep generative models, dethroning the longstanding supremacy of Generative Adversarial Networks (GANs) in the demanding domain of image generation. They are, essentially, a group of probabilistic generative models that begin by gradually introducing/diffusing noise into the training data. Through this iterative process, the models learn to reverse the diffusion chain, allowing them to generate entirely new samples out of pure noise.

In this project, we will experiment hands-on with popular frameworks of Text-to-Image diffusion models. Like any tool, these models have their strengths and limitations, which we aim to uncover. The primary goals of this project are to gain hands-on experience with state-of-the-art pretrained image generative models and explore the strengths and weaknesses of these models by crafting diverse and thought-provoking prompts. While Text-to-Image diffusion models perform well when prompts are clear and simplistic, their behaviour in response to ambiguous or confusing prompts, such as “a fiery ocean of ice” or “paint a square circle” is unpredictable. In essence, through prompting, we will analyse model responses to different types of input to uncover their capabilities and limitations. Additionally, we will compare different publicly available frameworks to explore their differences.

Απαιτούμενες/επιθυμητές γνώσεις: Εξοικείωση με Python και machine learning frameworks (PyTorch, TensorFlow). Στοιχειώδεις γνώσεις σε Computer Vision. Για περισσότερες πληροφορίες επικοινωνήστε με τον Θ. Βουλόδημο (τηλ. 210-7723040, e-mail: thanosv at mail.ntua.gr), την Παρασκευή Τζούβελη (e-mail: tpar at image.ece.ntua.gr) και τον Λευτέρη Τσώνη (e-mail: lefteristsonis at outlook.com)

[1] Yang, L. and Zhang, Z. et al. (2022) ‘Diffusion Models: A Comprehensive Survey of Methods and Applications’

[2] Rombach R. and Blattmann A. et al. (2022) ‘High-Resolution Image Synthesis with Latent Diffusion Models’ [CVPR ‘22]

[3] Saharia C. and Chan W. et al. (2022) ‘Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding’

Prototype Explanations

Εργασία Μαθήματος 6

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	Explainability
Τύπος Εργασίας	Survey

Περιγραφή

Despite the rapid evolution of XAI techniques, critiques have been made regarding the human comprehensibility of certain methods. Beyond instance-based local techniques, there exist global explainability approaches that aim to impart readers with a more comprehensive grasp of the overarching patterns that define data clusters within a specific classifier's perspective. Prototype-based explanations leverage representative examples or prototypes to provide intuitive insights into the decision-making processes of machine learning models. Global exemplars yield localized explanations, resulting in easily comprehensible "This Looks Like That, Because…" explanations. In this comprehensive exploration of prototype explanations, we will investigate their versatility across various domains, modalities, and a wide spectrum of machine learning models, including CNNs and GNNs. By categorizing specific use cases and delving into prototype generation methods, visualization techniques, and evaluation approaches, we will uncover valuable insights into the practicality of this explanation approach. Additionally, we will identify future directions for enhancing XAI applications while addressing associated challenges.

Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr) και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr).

Word Sense Disambiguation Using GNNs

Εργασία Μαθήματος 7

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	GNNs, LLMs, Application in Medical Domain
Τύπος Εργασίας	Reproduction

Περιγραφή

Graph Neural Networks (GNNs) have been increasingly gaining popularity across many fields, including of course NLP. Within the realm of NLP, Word Sense Disambiguation (WSD) holds significant importance, particularly in the context of emerging intelligent search and retrieval technologies, such as chatbots. These technologies demand robust approaches for grasping contextual nuances. This assignment entails replicating the experiments conducted in Vretinaris et al.'s [1] study. Their paper offers a compelling introduction to WSD through GNNs, making the topic accessible even to those with a foundational understanding. It provides a practical illustration of GNNs' real-world applicability through an application to the medical domain. It further delves into the intricate details of a GNN-based method for medical entity disambiguation, elucidating the optimization strategies applied for method refinement. Throughout its narrative, the paper adeptly demystifies key concepts and techniques related to GNNs, offering clarity to readers with a basic GNN background. Furthermore, the comprehensive experimental assessment of the proposed approach equips readers with practical insights into GNN utilization, bolstering their ability to apply this technology in real-world contexts. Upon finishing the replication process, the assignee will have acquired expertise in WSD within a critical domain, a more profound comprehension of heterogeneous GNNs, and proficiency in the development and optimization of such algorithms. This foundation will enable the pursuit of future endeavors, such as exploring alternative negative sampling methods or adapting these techniques to different tasks like entity linking or relation prediction.

[1] https://dl.acm.org/doi/pdf/10.1145/3448016.3457328

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, εξοικείωση με pytorch και PyG. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr) και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr).

GNNs for Question Answering

Εργασία Μαθήματος 8

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	GNNs, LLMs
Τύπος Εργασίας	Survey

Περιγραφή

In recent years, there has been an explosive surge in research papers related to Graph Neural Networks (GNNs) presented at top-tier AI conferences worldwide. These papers encompass novel GNN techniques and optimizations, as well as GNNs serving as tools across various applications. Consequently, there is a growing need for survey papers that systematically categorize and consolidate GNN-based methods within specific domains. These surveys are essential for summarizing the extensive wealth of information and insights, providing a comprehensive overview of the evolving GNN landscape. Question Answering (QA) involves understanding and generating responses from natural language queries, often requiring complex knowledge representation and reasoning. GNNs, with their ability to model structured data and relationships, offer a promising avenue for enhancing QA performance. Conducting a survey that delves into the intersection of GNNs and QA would be of paramount importance. Such a survey would consolidate the diverse approaches and applications of GNNs in QA, provide insights into benchmarking and evaluation, highlight real-world use cases, and offer guidance to researchers and practitioners, ultimately fostering advancements in this critical area of natural language understanding.

Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr).

Confidence Calibration for Graph Neural Networks

Εργασία Μαθήματος 9

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	GNNs
Τύπος Εργασίας	Reproduction

Περιγραφή

Graph Neural Networks (GNNs) are a type of neural network that can effectively learn node representations based on the message-passing manner, making them well-suited for dealing with graph data. GNNs have been applied to various applications, including node classification, link prediction, and graph classification, and have achieved remarkable accuracy. However, despite their success, the trustworthiness of GNNs is still unexplored. Previous studies suggest that many modern neural networks are over-confident on their predictions, but surprisingly, GNNs are primarily under-confident. Therefore, confidence calibration for GNNs is highly desired. Confidence calibration aims to improve the reliability of the model's predictions by adjusting the confidence scores assigned to each prediction. In this way, calibrated GNNs can provide more trustworthy predictions, which is crucial for real-world applications, especially in safety-critical fields. This assignment involves reproducing the results of Wang et al. [1]. They propose a novel, post-hoc method, to calibrate GNNs, by training another GNN (CaGCN), whose sole purpose is to calibrate the results of the main model. This is the first endeavor, at employing a neural network to calibrate the confidence of the model, without using any of the more conventional calibration techniques. A variety of datasets are utilized to evaluate the proposed method, including Cora, CiteSeer, Pubmed and CoraFull. Following the conclusion of the assignment, a substantial research landscape emerges with regard to this topic. This landscape encompasses endeavors such as the optimization and enhancement of the presented architectural framework, alongside comprehensive evaluations across diverse datasets and tasks.

[1] https://proceedings.neurips.cc/paper/2021/hash/c7a9f13a6c0940277d46706c7ca32601-Abstract.html

Απαιτούμενες/επιθυμητές γνώσεις: Python, θεωρητικό υπόβαθρο και μεθοδολογίες για γράφους, εξοικείωση με pytorch και PyG. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr), το Γιώργο Φιλανδριανό (geofila@islab.ntua.gr) και την Αγγελική Δημητρίου (angelikidim@islab.ntua.gr).

Graph Autoencoders for Unsupervised Machine Learning with Graphs

Εργασία Μαθήματος 10

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	GNNs
Τύπος Εργασίας	Survey

Περιγραφή

Autoencoders, a fundamental concept in machine learning, are neural networks designed for unsupervised learning tasks, primarily used in dimensionality reduction and feature learning. They consist of an encoder network, which compresses input data into a lower-dimensional representation, and a decoder network that reconstructs the original input from this representation. Graph Autoencoders, an extension of traditional autoencoders, have emerged as a vital tool in the realm of graph data analysis. They play a crucial role in learning meaningful representations of graph-structured data, such as social networks, recommendation systems, and biological networks. Graph Autoencoders encode graph nodes and edges into a lower-dimensional space, allowing for various downstream tasks like link prediction, node classification, and community detection. With the rapid emergence of GNNs across most conferences and journals, there has been a substantial amount of research on Graph Autoencoders. So far, most GNN surveys, only briefly mention how Graph Autoencoders work, or they analyze them for a specific application (e.g. Graph Generation). For this reason, this thesis will focus on writing a comprehensive survey of the several architectures and learning-methods proposed, something that would be helpful for the community, as well as practical for future reference.

Απαιτούμενες/επιθυμητές γνώσεις: Python, θεωρητικό υπόβαθρο και μεθοδολογίες για γράφους, εξοικείωση με pytorch και PyG. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr).

Narration to Navigation: Creating a Challenging Dataset for Text to Knowledge Graph Creation

Εργασία Μαθήματος 11

Επιβλέπων	Γιώργος Στάμου
Σχετιζόμενα μαθήματα	Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα και Βαθιά Μάθηση
Status	Διαθέσιμη
Περιοχή	GNNs, LLMs
Τύπος Εργασίας	Dataset

Περιγραφή

While clear factual data, like Wikipedia articles, have been extensively explored (as demonstrated by projects like WebNLG [1, 2]), the creation of a dataset containing text and corresponding knowledge graphs from more diverse and unstructured sources presents a unique opportunity. Narration-type texts, in particular, introduce a myriad of challenges and complexities that make them intriguing candidates for knowledge extraction. These texts exhibit temporal dependencies, convey nuanced negations, and represent language in a real-world context that goes beyond the realm of straightforward facts and figures. This dataset's potential sources are boundless, ranging from news articles to academic papers, books, and even personal diary entries. By providing a structured representation of language in the wild, this dataset would enable the development of advanced LLM prompting techniques or GNN methods capable of extracting such knowledge from a broader spectrum of contexts. In the context of this assignment, the main objective is the creation of such a dataset, beginning with the initial step of identifying suitable source corpora. Established datasets like ROCStories, BookSum, AG News, or data obtained from online platforms like blogs and arXiv can serve as foundational resources to initiate the process. Then, employing a combination of standardized techniques and human intelligence, including potential manual extraction processes, to extract knowledge from these sources is the final and most crucial step.

[1]https://synalp.gitlabpages.inria.fr/webnlg-challenge/
[2] https://aclanthology.org/2022.findings-emnlp.116.pdf
[3] https://arxiv.org/pdf/2309.11669.pdf

Απαιτούμενες/επιθυμητές γνώσεις: Python, τεχνικές επεξεργασίας φυσικής γλώσσας, θεωρητικό υπόβαθρο και μεθοδολογίες για γράφους γνώσης. Για περισσότερες πληροφορίες επικοινωνήστε με τον Γ. Στάμου (τηλ. 210-7723040, e-mail: gstam at cs.ntua.gr), Αγγελική Δημητρίου (angelikidim@islab.ntua.gr) και τη Μαρία Λυμπεραίου (E-mail: marialymp@islab.ntua.gr).