UDC 004.89
DOI: https://doi.org/10.17721/1812-5409.2024/1.25

Ruslan PRAVOSUD, PhD Student
ORCID ID: 0009-0008-9450-5515
e-mail: pronod9999@gmail.com
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

SURVEY ON COMBINATION OF NATURAL LANGUAGE PROCESSING AND REINFORCEMENT LEARNING ALGORITHMS

The integration of NLP and RL has gained significant attention in recent years, as it holds the potential to enhance the capabilities of various applications, ranging from language understanding and generation to dialogue systems and autonomous agents. The incorporation of RL into NLP algorithms enhances language-related tasks by enabling adaptation and learning from interactions and feedback. This integration proves valuable in scenarios where language understanding and generation require dynamic and context-dependent responses, contributing to improved real-world performance.

The following methods were used to investigate the algorithms combining NLP and RL: Analysis, comparison.

NLP and RL algorithms combination provides state of the art results in a lot of AI-related tasks' solutions. This topic is a promising area of research for NLP based applications, which will probably lead to considerable improvements in quality of existing ones.

The survey explores the challenges and opportunities in fusing NLP and RL. Furthermore, it investigates the impact of different RL paradigms applications on NLP algorithms performance and combination of NLP and RL in more complex systems like simulated or real world navigation, which also includes usage of Computer Vision subsystems. In addition to reviewing existing research results, the paper identifies potential avenues for future research and development in the field.

Keywords: *Natural Language Processing, Reinforcement Learning, Robotics, Computer Vision.*

Introduction

Natural Language Processing (NLP) is a branch of artificial intelligence that deals with the interaction between computers and human language. The primary goal of NLP is to enable machines to understand, interpret, and generate human-like text or speech. This interdisciplinary field combines computer science, linguistics, and cognitive psychology to develop algorithms and models that can process, analyze, and derive meaning from vast amounts of natural language data. NLP applications range from language translation and sentiment analysis to chatbots and voice recognition systems, contributing to the development of more intuitive and effective human-computer interfaces.

Reinforcement Learning (RL) is a subfield of machine learning that focuses on training agents to make sequential decisions by interacting with an environment (Sutton, & Barto, 2018). The learning process involves the agent receiving feedback in the form of rewards or penalties based on its actions, guiding it to discover optimal strategies over time. Through trial and error, the agent learns to maximize cumulative rewards by adapting its behavior in response to environmental feedback. RL has found applications in various domains, including robotics, gaming, and autonomous systems, where the agent learns to navigate complex environments and make decisions that lead to favorable outcomes.

The integration of reinforcement learning into natural language processing algorithms offers promising advancements in several key language-related tasks (Uc-Cetina et al., 2023). In machine translation, RL can enhance training by allowing models to adapt translations based on dynamic language contexts, improving overall accuracy. Dialogue systems, like chatbots, benefit from RL in optimizing responses to user input, resulting in more contextually relevant and engaging conversations. RL is particularly useful in text summarization, where it helps generate concise and informative summaries by incorporating reward signals for the relevance and informativeness of the output.

Sentiment analysis tasks can be optimized using RL, allowing models to adapt to different contexts and preferences, leading to more accurate and personalized sentiment predictions. Named Entity Recognition (NER) systems can also benefit from RL by optimizing the extraction of entities in diverse contexts, improving adaptability to various domains. In information retrieval, RL improves search engine performance by learning to rank results based on user feedback, providing more relevant and personalized search outcomes. Additionally, RL aids text generation tasks, guiding models to produce coherent, contextually relevant, and engaging content, overcoming challenges in generating diverse and appropriate responses.

Furthermore, a lot of interesting applications of combination NLP and RL algorithms are focused on integrating both of them into complex systems consisting of several interdomains parts. In this paper we describe such application to navigation in real and simulated environments.

Main results and methods used

RL applications for NLP tasks solutions improvement. Reinforcement Learning applications in Natural Language Processing tasks offer promising avenues for improving solution effectiveness. By integrating RL into NLP algorithms, models can adapt and enhance performance through learning from interactions and feedback. In machine translation, RL enables models to dynamically adjust translations based on contextual nuances, improving overall accuracy. Dialogue systems benefit by optimizing responses through RL, resulting in more contextually relevant and engaging conversations. RL proves valuable in text summarization, guiding models to generate concise and informative summaries by considering reward signals for relevance and informativeness. Sentiment analysis, named entity recognition, information retrieval, and text generation tasks all see improvements when RL is applied, allowing models to adapt to different contexts, optimize entity extraction, enhance search engine results, and generate more coherent and contextually relevant content. In essence, RL applications empower NLP solutions to learn, adapt, and excel in dynamic language-related scenarios.

Abstractive summary. Attentional, RNN-based encoder-decoder models have demonstrated effective performance in abstractive summarization, particularly on short input and output sequences. However, when applied to longer documents and summaries, these models often exhibit issues such as repetitive and incoherent phrases. A neural network model introduced by (Romain, Xiong, & Socher, 2017) study features an intra-attention mechanism that focuses on the input and continuously generated output separately. Additionally, a novel training approach is proposed, combining standard supervised word prediction with reinforcement learning.

Models trained solely through supervised learning frequently suffer from "exposure bias," assuming that ground truth is provided at each training step. The integration of standard word prediction with RL's global sequence prediction training mitigates this bias, resulting in more readable summaries. The evaluation of this model on the CNN/Daily Mail and New York Times datasets reveals significant improvements. Specifically, the model achieves a ROUGE-1 score of 41,16 on the CNN/Daily Mail dataset, surpassing the performance of previous state-of-the-art models. Human evaluation further confirms that the proposed model generates higher-quality summaries. In the nutshell, the RL incorporation downs to modifying of loss function add a RL reward. Optimization is performed by self-critical policy gradient training algorithm. It worth noting that contemporary plain NLP SOTA models surpassed the results of the paper. However, the analysis provided in the study shows that incorporating RL into vanilla ML model significantly improves the performance. Therefore, its incorporation into SOTA plain NLP models may improve current performance.

Question answering. In the domain of question answering, conventional models that are optimized with cross-entropy loss tend to prioritize precise answers, unintentionally penalizing nearby or overlapping responses that might be equally accurate. To overcome this drawback, the approach proposed by (Caiming, Zhong, & Socher, 2017) introduces a blended objective, combining cross-entropy loss with self-critical policy learning. This combined objective utilizes rewards derived from word overlap, effectively addressing the misalignment between the evaluation metric and the optimization goal.

Alongside the mixed objective, dynamic coattention networks (DCN) are enhanced by integrating a deep residual coattention encoder inspired by recent advancements in deep self-attention and residual networks. These enhancements significantly contribute to elevating model performance across diverse question types and input lengths, particularly improving the handling of lengthy questions that require capturing long-term dependencies.

The effectiveness of the model is validated on the Stanford Question Answering Dataset, achieving state-of-the-art results with an exact match accuracy of 75,1 % and an F1 score of 83,1 %. Moreover, the ensemble model demonstrates even superior performance, securing an exact match accuracy of 78,9 % and an F1 score of 86,0 %. This highlights the efficacy of the blended objective and the enhanced DCN architecture in advancing the capabilities of question answering systems.

In this study RL technique is used to blend together regural loss function and self-critical reinforcement learning objective. Ablation study was conducted that shows that incorporation of RL objective is considerably improving the overall model's performance.

Dialogue Generation. Advancements in neural models for dialogue generation hold significant potential in generating responses for conversational agents. However, these models often exhibit a short-sighted approach, predicting utterances one at a time without considering their impact on future outcomes. Recognizing the importance of modeling the future direction of a dialogue for generating coherent and engaging conversations, traditional Natural Language Processing models turned to reinforcement learning.

Renowned scientists present introduces a novel approach that integrates the objectives of generating coherent dialogues and considering future outcomes by applying deep reinforcement learning to model future rewards in chatbot dialogue (Li et al., 2016). The proposed model simulates dialogues between virtual agents, employing policy gradient methods to reward sequences that exhibit three essential conversational properties: informativeness, coherence, and ease of answering (related to forward-looking function). Evaluation metrics include diversity, length, and human judgment, demonstrating that the algorithm generates more interactive responses and facilitates a more sustained conversation in dialogue simulation.

This research represents an initial step towards developing a neural conversational model that takes into account the long-term success of dialogues. By addressing the limitations of short-sightedness in dialogue generation, the proposed approach contributes to the advancement of more effective and coherent conversational agents.

The model uses the encoder-decoder architecture as its backbone, and simulates conversation between two virtual agents to explore the space of possible actions while learning to maximize expected reward (Mnih et al., 2013; Xiong, Zhong, & Socher, 2016; Goodfellow et al., 2013; Srivastava, Greff, & Schmidhuber, 2016). The agent learns a policy by optimizing the long-term developer-defined reward from ongoing dialogue simulations using policy gradient methods. The result shows that RL based model achieves better scores then previous SOTA plain NLP ones.

RL applications to NLP algorithms summary. RL based NLP models had achieved better performance than plain NLP SOTA at the times of the papers 'publication. Currently, GPT (Brown et al., 2020; Radford et al., 2018) model surpasses all the previous ones in terms of performance and usability of NLP-based tools. However, it might also be improved by incorporating some kind of RL techniques. This can be a topic for future researches.

NLP and RL used together as parts of complex systems. NLP and RL play significant roles in advancing robotics, enabling more intuitive human-robot inter- actions and enhancing the autonomy of robotic systems (Anisimov, Marchenko, & Zemlianskyi, 2019; Anisimov, Marchenko, & Zemlianskyi, 2021).

In the context of NLP, robots equipped with language processing capabilities can understand and respond to natural language commands, making them more accessible to non-experts. This facilitates seamless communication between humans and robots, allowing users to issue instructions or queries in a manner akin to interacting with another person. NLP is employed in tasks such as voice commands, language-based navigation, and human-robot collaboration.

On the other hand, RL is crucial for endowing robots with the ability to learn and adapt in dynamic environments (Ammanabrolu et al., 2021; Colas, et al., 2020). RL algorithms enable robots to make decisions based on trial and error, learning optimal strategies through interactions with their surroundings. This is particularly valuable in scenarios where robots need to perform tasks without explicit programming, such as autonomous navigation, grasping objects, or manipulating their environment.

The combination of NLP and RL in robotics allows robots to understand natural language commands, learn from interactions, and adapt their behavior accordingly. This integration enhances the overall versatility, usability, and autonomy of robotic systems, paving the way for more sophisticated and user- friendly applications in fields such as assistive robotics, home automation, and industrial automation.

Robotic system trained to follow natural language directions. In the context of human-robot interaction, natural language serves as an instinctive and adaptable way for people to communicate with the robots, which are increasingly present in homes and workplaces. Despite recent advancements that enable robots to understand natural language manipulation and navigation commands, existing methods rely on having a pre-established map of the robot's environment. The paper introduces an innovative learning framework designed to empower robots to adeptly follow natural language route directions without prior knowledge of their surroundings.

The proposed algorithm (Hemachandra et al., 2015) leverages spatial and semantic information conveyed by humans through commands to acquire a distribution over the metric and semantic properties of spatially extended environments. This distribution replaces the need for a pre-existing world model, interpreting the natural language instruction as a distribution over intended behavior. The method employs a unique space planner that directly reasons over a map of the environment and behavior distributions, learning the policy through imitation learning. The framework is assessed on a voice-commandable wheelchair, illustrating that, by learning and making inferences over a latent environment model, the algorithm successfully follows natural language route directions within unfamiliar and extended environments. This research represents a significant advancement in enabling robots to comprehend and act upon natural language commands in diverse and unexplored settings.

Recent strides in vision and language methodologies have made remarkable advancements in closely related domains. This becomes particularly noteworthy as a robot interpreting a natural-language navigation command based on visual inputs engages in a vision and language process akin to Visual Question Answering. Both tasks can be seen as sequence-to-sequence translation challenges grounded in visual information, and many analogous methods can be applied.

To foster the application of vision and language methodologies to the task of interpreting visually-grounded navigation instructions, paper provides a solution is presented in the form of the Matterport3D Simulator. Unlike this approach processed images of surroundings as a part of the model, having both text input and environment image as a part of input for the RL agent that chooses the next action based on them. The simulator, rooted in real imagery and designed for large-scale reinforcement learning, serves as an environment for a variety of embodied vision and language tasks.

Grounded Language Learning. Grounded language acquisition focuses on the process of learning the meaning of language in the context of the physical world (Matuszek, 2018). With the growing capabilities and prevalence of robots, there is a rising demand for individuals without specialized expertise to engage with and command these robots. Natural language emerges as an instinctive, adaptable, and customizable means for communication in this context. Concurrently, physically embodied agents provide a pathway to comprehend natural language within the context of the real-world scenarios to which it pertains. This paper provides a comprehensive overview of the research domain, highlighting recent advancements, and outlines future directions and challenges that persist in this field.

NLP and RL applied together in robotics summary. There are a lot of researches focused on improving robotics performance by blending of RL, NLP, CV (Anderson et al., 2017; Rennie et al., 2017) and lots of other fields in difference aspects such as making a model to understand visual objects and textual description correspondence and controlling a robotic system through natural language commands for navigation in unknown environments. Some of them are focused on improvement of specific aspects of algorithms in real world, while others try them in simulation. Further researches are to be focused on blending all the techniques to create systems capable to be interacted with by free form Natural Language commands for managing robots navigating in real world environment or executing some other kinds of actions. Improved version of paper's algorithm, where an agent is trained to execute its own commands and come up with its own goals

to explore the possible actions based on NLP+RL framework, might be used to train such agents to understand free form Natural Language instructions and execute the desired action. While the proposed algorithm trains an agent with a strict simple grammar, it can be improved to perform arbitrary natural language commands on extended actions space.

Discussion and conclusions

Reinforcement learning has proven its beneficial impact on NLP algorithms. While current SOTA in NLP use plain Large Language Models, further researches are worth being conducted to investigate possible improvements by RL introduction, as it led to a significant raise in performance of their ancestors.

At the same time, NLP and RL create a symbiotic relationship in robotics. NLP provides a means for humans to convey instructions to robots in a familiar and accessible manner, while RL enables robots to continuously improve their decision-making processes and actions based on the context and feedback received. This combination is especially impactful in applications such as autonomous navigation, task execution, and human-robot collaboration, contributing to the development of more sophisticated and adaptable robotic systems.

References

Ammanabrolu, P., Urbanek, J., Li, M., Szlam, A., Rocktäschel, T., & Weston, J. (2021). *How to motivate your dragon: Teaching goal-driven agents to speak and act in fantasy worlds.* https://arxiv.org/abs/2010.00685

Anderson, P., Wu, Q., Teney, D., Bruce, J., Johnson, M., Sünderhauf, N., Reid, I., Gould, S., & Hengel, A. van den. (2018). Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. *Proceedings of the IEEE conference on computer vision and pattern recognition*. https://par.nsf.gov/servlets/purl/10066404

Anisimov, A. V., Marchenko, O. O. & Zemlianskyi, V. R. (2019). Evolutionary Method of Constructing Artificial Intelligence Systems. *Cybern Syst Anal,* 55, 1–9. https://doi.org/10.1007/s10559-019-00106-x

Anisimov, A. V., Marchenko, A. A., & Zemlianskyi, V. R. (2021). Influence of Language on the Lifespan of Populations of Artificial Intelligence. *Cybern Syst Anal,* 57, 669–675. https://doi.org/10.1007/s10559-021-00392-4

Brown, T. B., et al. (2020). *Language models are few-shot learners*. https://arxiv.org/abs/2005.14165

Colas, C., Karch, T., Lair, N., Dussoux, J.- M., Moulin-Frier, C., Dominey, P. F., & Oudeyer, P.- Y. (2020). Language as a cognitive tool to imagine goals in curiosity driven exploration. *Advances in Neural Information Processing Systems,* 33, 3761–3774. https://arxiv.org/abs/1711.07280

Hemachandra, S., Duvallet, F., Howard, Th. M., Roy, N., Stentz, A., & Walter, M. R. (2015). *Learning models for following natural language directions in unknown environments*. https://arxiv.org/pdf/1503.05079.pdf

Goodfellow, I. J., Warde-Farley, D., Mirza, M. Courville, A., & Bengio, Yo. (2013). Maxout networks. *International conference on machine learning*. PMLR. https://arxiv.org/abs/1302.4389

Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., & Jurafsky, D. (2016). *Deep reinforcement learning for dialogue generation*. https://arxiv.org/abs/1606.01541

Matuzek, C. (2018). *Grounded language learning: Where robotics and nlp meet (invited talk).* https://www.aclweb.org/anthology/D18-1355.pdf

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). *Playing atari with deep reinforcement learning*. https://arxiv.org/abs/1312.5602

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. https://s3-us-west-2.amazonaws.com /openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J., & Goel, V. (2017). Self-critical sequence training for image captioning. *Proceedings of the IEEE conference on computer vision and pattern recognition.* https://arxiv.org/abs/1612.00563

Romain, P., Xiong, C., & Socher, R., (2017). *A deep reinforced model for abstractive summarization.* https://arxiv.org/abs/1705.04304

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Training very deep networks. *Advances in neural information processing systems,* 28. https://arxiv.org/abs/1507.06228

Sutton, R., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press. https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

Uc-Cetina, V., Navarro-Guerrero, N., Martin-Gonzalez, A., Weber, C., & Wermter, S. (2023). Survey on reinforcement learning for language processing. *Artificial Intelligence Review,* 56(2), 1543–1575. https://arxiv.org/abs/2104.05565

Xiong, C., Zhong, V., & Socher. (2016). *Dynamic coattention networks for question answering*. https://arxiv.org/abs/1611.01604

Xiong, C., Zhong, V., & Socher, R. (2017). *Dcn+: Mixed objective and deep residual coattention for question answering*. https://arxiv.org/abs/1711.00106

Отримано редакцією журналу / Received: 18.01.24
Прорецензовано / Revised: 29.04.24
Схвалено до друку / Accepted: 30.04.24

Руслан ПРАВОСУД, асп.
ORCID ID: 0009-0008-9450-5515
e-mail: pronod9999@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

ОГЛЯД ПОЄДНАННЯ АЛГОРИТМІВ ОБРОБКИ ПРИРОДНИХ МОВ ТА НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Інтеграція обробки природної мови та навчання з підкріпленням останнім часом є актуальним предметом дослідження, оскільки вона має потенціал для підвищення якості роботи різних застосувань, починаючи з розуміння та генерації мови, і закінчуючи діалоговими системами й автономними агентами. Залучення RL в алгоритми NLP поліпшує виконання завдань, пов'язаних з мовою, даючи змогу адаптуватися і навчатися від взаємодій та зворотного зв'язку. Ця інтеграція виявляється цінною у сценаріях, де розуміння та генерація мови потребують динамічних і контекстно-залежних відповідей, що сприяє підвищенню реальної продуктивності. Для дослідження особливостей поєднання алгоритмів обробки природних мов і навчання з підкріпленням було застосовано такі методики, як аналіз та порівняння. Поєднання алгоритмів обробки природної мови та навчання з підкріпленням забезпечує найсучасніші результати у виконанні багатьох завдань, пов'язаних зі штучним інтелектом. Ця тема є перспективною галуззю досліджень для застосувань на основі NLP, що, імовірно, призведе до значного поліпшення якості наявних рішень.

У пропонованій статті досліджено виклики та можливості в поєднанні обробки природної мови та навчання з підкріпленням. Крім того, проаналізовано вплив застосування різних парадигм RL на продуктивність алгоритмів NLP та поєднання NLP і RL у складніших системах, таких як навігація у симульованому або реальному світі, що також передбачає використання підсистем комп'ютерного зору. Окрім огляду наявних результатів досліджень, у роботі визначено потенційні напрямки для майбутніх досліджень і розробок у цій галузі.

Ключові слова: ***обробка природної мови, навчання з підкріпленням, робототехніка, комп'ютерний зір.***

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.