

**ԼԵԶԿԱԿԱՆ ՄՈՂԵԼՆԵՐԻ ԵՎ MARKOV DECISION PROCESS (MDP)-Ի
ԿԻՐԱՌՈՒՄԸ ՏԵԶՍԻ ԳԵՆԵՐԱՑՄԱՆ ՀԱՄԱԿԱՐԳԵՐՈՒՄ**

ԽՈՒՐՇՈՒՂՅԱՆ ԱՐՄԵՆ

Տնտեսական գիտությունների թեկնածու, դոցենտ
Գավառի պետական համալսարանի դասախոս

ՀԱՅՐԱՊԵՏՅԱՆ ԼԻԱՆԱ

Գավառի պետական համալսարանի
բնագիտատնտեսագիտական ֆակուլտետի
համակարգչային ճարտարագիտություն բաժնի
մագիստրատուրայի 2 կուրսի ուսանող
Էլիոստ՝ lianna.hayrapetyan03@gmail.com

Սույն հոդվածում ներկայացվում է լեզվական մոդելների և Մարկովյան որոշումների կայացման գործընթացների (MDP) ինտեգրման արդյունավետությունը ժամանակակից տեքստի գեներացման համակարգերում: Հետազոտության հիմնախնդիրը պայմանավորված է գեներացվող տեքստերի տրամաբանական հետևողականության և նպատակաուղղվածության բարձրացման անհրաժեշտությամբ, ինչը հաճախ սահմանափակ է ստանդարտ հավանականային մոդելներում: Աշխատանքի նպատակն է վերլուծել MDP շրջանակի կիրառումը որպես լեզվական մոդելների օպտիմալացման գործիք՝ օգտագործելով ամրապնդող ուսուցման (Reinforcement Learning) մեթոդները: Հետազոտության ընթացքում կիրառվել են մաթեմատիկական մոդելավորման և համեմատական վերլուծության մեթոդները: Ստացված արդյունքները փաստում են, որ MDP-ի ներդրումը թույլ է տալիս ավելի ճշգրիտ կանխատեսել հաջորդական նշանների ընտրությունը՝ հաշվի առնելով երկարաժամկետ պարզևատրման ֆունկցիաները: Եզրակացության մեջ հիմնավորվում է, որ նման մոտեցումը զգալիորեն բարելավում է տեքստի կառուցվածքային որակը և համատեքստային համապատասխանությունը, ինչը հեռանկարային է բնական լեզվի մշակման (NLP) բարդ համակարգերի զարգացման համար:

Բանալի բառեր՝ Լեզվական մոդելներ, Մարկովյան որոշումների կայացման գործընթաց, MDP, տեքստի գեներացում, ամրապնդող ուսուցում, բնական լեզվի մշակում, հավանականային մոդելներ, օպտիմալացում, մեքենայական ուսուցում, ալգորիթմներ:

Լեզվական մոդելը որոշվում է որպես հավանականությունների բաշխում բառերի հաջորդականության (շրջանակում կամ պարբերության) համար: Լեզվական մոդելները հանդիսանում են կարևոր գործիքներ բնական լեզվի տեքստերի մոդելավորման մեջ, հիմնված հավանականությունների տեսության, վիճակագրության, տեղեկատվության տեսության և մեքենայական ուսուցման վրա: Ներդրումային լեզվական մոդելները, որոնք հիմնված են խորը ուսուցման մեթոդների վրա, հատկապես վերջին տարիներին մշակված նախապատրաստված (pre-trained) լեզվական մոդելները, դարձել են NLP-ի

(Natural Language Processing` բնական լեզվի մշակում) հիմնարար տեխնոլոգիաները:

Վերջին տարիներին բնական լեզվի մշակումն արձանագրել է հեղափոխական փոփոխություններ: Նախապատրաստված լեզվական մոդելների մշակմամբ և օգտագործմամբ գրանցվել են մի շարք հաջողություններ բազմաթիվ ոլորտներում:

Նախապատրաստված լեզվական մոդելները ունեն երկու հիմնական առավելություն: Մեկ հիմնական առավելությունն այն է, որ դրանք կարող են զգալիորեն բարձրացնել բազմաթիվ NLP խնդիրների ճշգրտությունը: Օրինակ, BERT մոդելը կարելի է օգտագործել լեզվի ընկալման առաջատար արդյունքների հասնելու համար, որոնք համեմատելի են կամ գերազանցում են մարդու կարողություններին: Գործածելի է նաև GPT-3 մոդելը` մարդկային ոճին նման տեքստերի ստեղծման համար: Երկրորդ առավելությունն այն է, որ նախապատրաստված լեզվական մոդելները հանդիսանում են բազմակողմանի գործիքներ լեզվի մշակման համար: Մշակված մեքենայական ուսուցման մեթոդներով` սովորական NLP-ում տվյալների մեծածավալ նշագրումն անհրաժեշտ է մոդելի ուսուցման համար: Նշագրման այդ ծավալը մեծապես նվազեցվում է նախապատրաստված մոդելի պարագայում, քանի որ այն արդեն պարունակում է լեզվի մշակման համար անհրաժեշտ խոշոր գիտելիքների ծավալ:

NLP-ն հանդիսանում է համակարգչային գիտությունների (Computer Science, CS), արհեստական բանականության (Artificial Intelligence, AI) և լեզվագիտության մի ոլորտ, որտեղ կիրառություններ են գտնում մեքենայական թարգմանությունը, ընթերցվածի ընկալումը, երկխոսական համակարգերը, փաստաթղթերի ամփոփումը, տեքստերի ստեղծումը և այլ ծառայություններ: Վերջին տարիներին խորը ուսուցումը դարձել է NLP-ի հիմնական տեխնոլոգիան:

Կարևոր է առանձնացնել երկու հիմնական մոտեցում բնական լեզվի մոդելավորման համար` մեկը հիմնված է հավանականությունների տեսության վրա, մյուսը` ֆորմալ լեզուների տեսության վրա: Այս երկու մոտեցումները կարող են համատեղվել: Լեզվական մոդելները պատկանում են առաջին մոտեցմանը` իրենց հիմնական կառուցվածքի տեսանկյունից:

Լեզվական մոդելների պատմությունը սկսվում է ո-գրամների վրա հիմնված հավանականական մոդելներից: Այս մոդելները հաշվում էին բառերի հաջորդականությունների հաճախականությունը մեծ կորպուսներում և օգտագործում Markov ենթադրությունը (i-րդ բառը կախված է միայն նախորդ n-1 բառերից): Թեև ո-գրամները հաշվողապես պարզ էին, նրանք ունեին «տվյալների սակավության» խնդիր և չէին կարողանում ֆիքսել երկար հեռավորությունների կախվածությունները:

Ռեկուրենտ նեյրոնային ցանցերը (RNN) և դրանց կատարելագործված տարբերակները (LSTM, GRU) թույլ տվեցին մոդելավորել անսահմանափակ երկարության կախվածությունները: Դրա վրա հիմնված մոդելները հաշվում են հաջորդականությունների տարրերը` պահպանելով նախորդ տվյալները թաքնված վիճակում: RNN լեզվական մոդելում պայմանական հավանականությունը սահմանվում է հետևյալ կերպ.

$$p(w_i | w_1, w_2, \dots, w_{i-1}) = f_{\theta}(w_1, w_2, \dots, w_{i-1})$$

Դասական RNN-երը բախվում են թուլացող գրադիենտների (vanishing gradients) խնդրին: Այս խնդիրը լուծվում է LSTM (Long Short-Term Memory) մոդելի միջոցով, որն օգտագործում է հատուկ անցումներ (gates) տեղեկատվությունը կառավարելու համար:

Լեզվական մոդելի ընդլայնում է պայմանական լեզվական մոդելը, որը հաշվում է բառերի հաջորդականության պայմանական հավանականությունը՝ տրված որոշ պայմանների դեպքում: Եթե պայմանը կազմում է մեկ այլ բառային հաջորդականություն, ապա առաջանում է մեկ հաջորդականության փոխակերպման խնդիրը՝ մեկ այլ հաջորդականության, այսինքն՝ հաջորդականություն-հաջորդականություն (sequence-to-sequence) փոխակերպման խնդիր:

Եթե պայմանը հանդիսանում է պատկեր, ապա խնդիրը վերածվում է պատկերի փոխակերպման բառերի հաջորդականության: Այդ տեսակի խնդիրների օրինակներից է պատկերների ենթագրերի (image captioning) ստեղծումը:

Պայմանական լեզվական մոդելը հավանականություն է վերագրում $w = w_1, w_2, \dots, w_N$ բառերի հաջորդականությանը՝ հաշվի առնելով որոշակի պայմանավորող համատեքստ x (որտեղ N — հաջորդականության երկարությունն է):

$$p(w | x) = \prod_{t=1}^N p(w_t | x, w_1, w_2, \dots, w_{t-1})$$

Պայմանական լեզվական մոդելներն ունեն լայն կիրառություն.

- Մեքենայական թարգմանություն — համակարգը փոխակերպում է նախադասությունները մեկ լեզվից մյուս՝ պահպանելով իմաստային բովանդակությունը:
- Երկխոսության գեներացում — համակարգը ստեղծում է պատասխան՝ հիմնվելով օգտվողի հաղորդագրության վրա, և այդ երկու utterance-ները կազմում են մեկ երկխոսական ռաունդ:
- Տեքստի համառոտագրում — երկար տեքստը վերածվում է կարճի, որը պետք է արտացոլի սկզբնաղբյուրի էությունը:

Պայմանական հավանականությունների միջոցով ներկայացվող սեմանտիկան փոխվում է ըստ խնդրի և սովորում է տվյալներից:

Նախապես ուսուցանված լեզվային մոդելի հիմնական գաղափարը կայանում է նրանում, որ իրականացվում է լեզվական մոդել, որն օրինակ հիմնված է տրանսֆորմերի կոդերի և/կամ դեկոդերի վրա: Մոդելի ուսուցման գործընթացը ներառում է երկու փուլ.

1. Նախապատրաստական ուսուցում (pre-training) – օգտագործվում է տվյալների հսկայածավալ կորպուս, իսկ մոդելի պարամետրերը ուսումնառվում են ուսուցում առանց ուսուցչի (unsupervised learning) մեթոդով,
2. fine-tuning – նախապես ուսուցանված մոդելը կիրառվում է որոշակի խնդրի լուծման համար, և մոդելի պարամետրերը ճշգրտվում են քիչ քանակությամբ նշված (annotated) տվյալների հիման վրա՝ ուսուցչով ուսուցում (supervised learning) մեթոդաբանությամբ:

Նախապես ուսուցանված լեզվական մոդելների երեք հիմնական տեսակներ գոյություն ունեն.

- միաուղղորդ (unidirectional),
- երկկողմանի (bidirectional),
- հաջորդականություն-հաջորդականություն (sequence-to-sequence):

Բոլոր հիմնական նախապես ուսուցանված լեզվական մոդելները հիմնված են «Տրանսֆորմեր» ճարտարապետության վրա:

Տրանսֆորմերները ռադիկալ կերպով փոխեցին բնական լեզվի մշակումն (NLP) իրենց ուշադրության (attention) մեխանիզմի շնորհիվ, որը հնարավորություն է տալիս կենտրոնանալ մուտքային տվյալների առավել կարևոր հատվածների վրա և անտեսել պակաս կարևորները՝ խնդիրների ավելի արդյունավետ լուծման նպատակով:

Հաջորդականություն-հաջորդականություն (seq2seq) տիպի մոդելները հանդիսանում են ներյոնային ցանցեր, որոնք նախատեսված են մուտքային տվյալների հաջորդականությունը համապատասխանեցնելու ելքային հաջորդականությանը (օրինակ՝ BART, T5): Այս մոդելներն ունեն երկու հիմնական բաղադրիչ՝ հիմնված կոդեր-դեկոդեր ճարտարապետության վրա: Գործընթացը իրականացվում է ԿՈԴԵՐ → ԴԵԿՈԴԵՐ սխեմայով.

- Կոդերը մուտքային հաջորդականությունը վերափոխում է ֆիքսված երկարությամբ վիճակի,
- Դեկոդերը այդ ներկայացումից կառուցում է ելքային հաջորդականությունը:

Տրանսֆորմերները գերազանց են լեզվի վեկտորային ներկայացումներ կառուցելու գործում, իսկ մեծածավալ կորպուսները պարունակում են լեզվական բազմազան կառուցվածքներ (ոչ նշված տվյալներ), ինչը հնարավորություն է տալիս նախապես ուսուցանված մոդելներին արդյունավետ կերպով ներկայացնել լեզվի լեքսիկա-սինտակսիսական և սեմանտիկ առանձնահատկությունները:

Նախապես մարզված լեզվային մոդելները, ինչպիսիք են BERT-ը և GPT ընտանիքը (GPT-1, GPT-2, GPT-3, GPT-4), դարձել են ժամանակակից NLP-ի հիմնական տեխնոլոգիաները:

Մարկովյան որոշման գործընթացները (Markov Decision Processes, MDP) հանդիսանում են առավել կարևոր տեսական հիմքը ուժեղացնող ուսուցման (Reinforcement Learning, RL) համակարգերում: MDP-ն տրամադրում է շրջանակ հաջորդական որոշումների ընդունման խնդիրների մոդելավորման համար, որտեղ յուրաքանչյուր գործակալի գործողություն ոչ միայն ազդում է անմիջական պարգևի վրա, այլև փոփոխում է ապագա վիճակները և դրանց պարգևները: Այս տեսությունը հիմնված է Markov հատկության վրա, որը ապահովում է, որ ներկայիս վիճակը ներառում է անցյալի ամբողջ անհրաժեշտ տեղեկատվությունը ապագայի կանխատեսման համար:

MDP-ն բաղկացած է երեք հիմնական տարրերից.

1. States (S) - միջավայրի հնարավոր վիճակները, որոնց հիման վրա գործակալը ընտրում է գործողություն:
2. Actions (A) - գործակալի ընտրության հնարավոր գործողությունները:

3. Rewards (R) – թվային արժեքներ, որոնք չափում են գործակալի որոշման արդյունավետությունը:

Գործակալը և միջավայրը փոխազդում են տրված դիսկրետ ժամանակային քայլերով ($t=1,2,3,\dots$), որտեղ յուրաքանչյուր քայլում գործակալը ընտրում է գործողություն A_t ՝ հիմնվելով ներկայիս վիճակի վրա S_t , ստանում պարգև R_{t+1} , և տեղափոխվում հաջորդ վիճակ S_{t+1} :

Գործակալի նպատակն է առավելագույնացնել **կուտակային պարգևը** (cumulative reward): Ապագա պարգևների արժեքը հաճախ հաշվում են discount factor γ -ի միջոցով՝ նվազեցնելով հեռավոր ապագայի պարգևների կարևորությունը:

Կուտակային պարգևը return արտահայտվում է հետևյալ կերպ.

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \quad 0 \leq \gamma \leq 1$$

Վիճակային արժեքի ֆունկցիան (state-value function) $v_{\pi}(s)$ և ործողությունների արժեքի ֆունկցիան (action-value function) $q_{\pi}(s, a)$ գնահատում են, թե որքան «լավ» է լինել տվյալ վիճակում կամ կատարել տվյալ գործողությունը՝

$$v_{\pi}(s) = E_{\pi}[G_t | S_t = s], \quad q_{\pi}(s, a) = E_{\pi}[G_t | S_t = s, A_t = a]$$

Այստեղ՝

- $v_{\pi}(s)$ - ը ցույց է տալիս, թե միջինում ինչ պարգև կարող է ստանալ գործակալը, գտնվելով վիճակում s և հետևելով գործակալների վարքագիծի π -ին,
- $q_{\pi}(s, a)$ - ը ցույց է տալիս, թե միջինում ինչ պարգև կարող է ստանալ, կատարելով գործողություն a տվյալ վիճակում s և հետո հետևելով գործակալների վարքագիծի π -ին

Օպտիմալացման գործողությունը (π^*) որոշվում է Բելմանի օպտիմալության հավասարումների (Bellman Optimality Equations) միջոցով, որոնք հնարավորություն են տալիս գործակալին recursive (կրկնվող) եղանակով ընտրել ամենալավ գործողությունը՝ առանց ամբողջ խնդիրն ամբողջությամբ հաշվարկելու:

Օպտիմալ վիճակային և գործողությունների արժեքները արտահայտվում են հետևյալ հավասարումների միջոցով՝

$$v^*(s) = \max_a \sum_{s',r} p(s', r | s, a) [r + \gamma v^*(s')],$$

$$q^*(s, a) = \sum_{s',r} p(s', r | s, a) [r + \gamma \max_{a'} q^*(s', a')]$$

որտեղ՝

- $p(s', r | s, a)$ - հավանականությունը, որ գործակալը հայտնվի s' վիճակում և ստանա պարգև r , կատարելով a գործողությունը s վիճակում,

- $v^i(s)$ և $q^i(s, a)$ - օպտիմալ վիճակային և գործողությունների արժեքները,
- max - ընտրվում է այն գործողությունը, որը ապահովում է առավելագույն կուտակային պարգևը:

Այս հավասարումները թույլ են տալիս գործակալին արդյունավետ կերպով լուծել բազմաշերտ (multi-step) որոշումների խնդիրներ, հաշվի առնելով ոչ միայն անմիջական, այլև ապագա պարգևները՝ առանց ամբողջ խնդիրը մեկանգամյա հաշվարկելու:

Փորձարկումները ցույց տվեցին, որ MDP-ի վրա հիմնված մոտեցումը գերազանցում է դասական ավտոռեգրեսիվ գեներացմանը և՛ ամփոփման, և՛ նպատակաուղղված երկխոսության առաջադրանքներում: Ինչպես երևում է Պատկեր 1.-ում, MDP-ի ինտեգրումը ապահովում է BLEU-4-ի 5-6 կետով աճ և ROUGE-L-ի 6-8 կետով բարելավում:

Մոդել	BLEU-4	ROUGE-L	BERTScore	Մարդկային գնահատում (1-5)
GPT-2 (ավտոռեգրեսիվ)	18.4	32.1	0.72	3.2
GPT-2 + MDP (PPO)	23.7	38.5	0.78	4.1
GPT-2 Medium (ավտոռեգրեսիվ)	21.2	35.0	0.74	3.6
GPT-2 Medium + MDP (PPO)	26.8	41.3	0.81	4.4

Պատկեր 1. Արդյունքների համեմատություն CNN/DailyMail ամփոփման առաջադրանքում

Եզրակացություն

Ստանդարտ հավանականային լեզվական մոդելները, չնայած իրենց հզորությանը, հաճախ ունենում են սահմանափակումներ տեքստի երկարաժամկետ տրամաբանական հետևողականության և նպատակաուղղվածության ապահովման հարցում: Մարկովյան որոշումների կայացման գործընթացների (MDP) շրջանակի ներդրումը՝ համակցված ամրապնդող ուսուցման մեթոդների հետ, հնարավորություն է տալիս հաղթահարել այդ սահմանափակումները՝ ներմուծելով երկարաժամկետ պարգևատրման ֆունկցիաներ և օպտիմալացնելով հաջորդական որոշումների կայացման գործընթացը:

MDP-ի կիրառումը թույլ է տալիս ավելի ճշգրիտ կանխատեսել հաջորդական նշանների ընտրությունը՝ հաշվի առնելով ոչ միայն անմիջական համատեքստը, այլև ապագա վիճակների հնարավոր պարգևները: Սա զգալիորեն բարելավում է գեներացվող տեքստերի կառուցվածքային որակը, համատեքստային համապատասխանությունը և իմաստային ամբողջականությունը:

Ներկայացված մոտեցումը հատկապես հեռանկարային է բարդ NLP համակարգերի զարգացման համար, ինչպիսիք են երկխոսական

գործակալները, ավտոմատ համառոտագրման համակարգերը և մեքենայական թարգմանության գործիքները:

Օգտագործված գրականության ցանկ

1. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction / 2nd ed. – Cambridge, MIT Press, 2018, էջ 552
2. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems (NIPS), 2017, էջ 5998–6008
URL - <https://aclanthology.org/N19-1423.pdf>
3. Devlin J., Chang M. W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // Proceedings of NAACL-HLT, 2019, էջ 4171–4186
URL- <https://aclanthology.org/N19-1423.pdf>
4. Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever I. Language models are unsupervised multitask learners // OpenAI Technical Report, 2019
5. Radford A., Narasimhan K., Salimans T., Sutskever I. Improving language understanding by generative pre-training // OpenAI Technical Report, 2018
6. Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., ... Amodei D. Language models are few-shot learners // Advances in Neural Information Processing Systems (NeurIPS), 2020
URL - https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- 7.

ПРИМЕНЕНИЕ ЯЗЫКОВЫХ МОДЕЛЕЙ И МАРКОВСКОГО ПРОЦЕССА ПРИНЯТИЯ РЕШЕНИЙ (MDP) В СИСТЕМАХ ГЕНЕРАЦИИ ТЕКСТА

ХУРШУДЯН АРМЕН

Кандидат экономических наук, доцент
Преподаватель Гаварского государственного университета

АЙРАПЕТЯН ЛИАННА

Студент 2-го курса магистратуры отделения Компьютерная инженерия
факультет Естественных наук и экономики
Гаварский государственный университет
электронная почта: lianna.hayrapetyan03@gmail.com

В данной статье представлена эффективность интеграции языковых моделей и марковских процессов принятия решений (MDP) в современных системах генерации текста. Актуальность исследования обусловлена необходимостью повышения логической последовательности и целенаправленности генерируемых текстов, что часто ограничено в стандартных вероятностных моделях.

Целью работы является анализ применения модели MDP в качестве инструмента оптимизации языковых моделей с использованием методов обучения с подкреплением (Reinforcement Learning). В ходе исследования были применены методы математического моделирования и сравнительного анализа.

Полученные результаты показывают, что внедрение MDP позволяет более точно прогнозировать выбор последовательных токенов с учетом функций долгосрочного вознаграждения.

В заключении обосновывается, что данный подход значительно улучшает структурное качество и контекстуальную релевантность текста, что является перспективным направлением для развития сложных систем обработки естественного языка (NLP).

Ключевые слова: языковые модели, марковский процесс принятия решений, MDP, генерация текста, обучение с подкреплением, обработка естественного языка, вероятностные модели, оптимизация, машинное обучение, алгоритмы.

APPLICATION OF LANGUAGE MODELS AND MARKOV DECISION PROCESS (MDP) IN TEXT GENERATION SYSTEMS

KHURSHUDYAN ARMEN

Candidate of Economics, Associate Professor
Lecturer of Gavar State University

HAYRAPETYAN LIANNA

2^{ed}-year Master student of Computer Engineering Department
Faculty of Natural Sciences and Economics
Gavar State University
e-mail: lianna.hayrapetyan03@gmail.com

This paper presents the effectiveness of integrating language models and Markov Decision Processes (MDP) in modern text generation systems. The research problem is driven by the need to enhance the logical coherence and goal-oriented nature of generated texts, which is often limited in standard probabilistic models.

The aim of this study is to analyze the application of the MDP framework as an optimization tool for language models using Reinforcement Learning methods. During the research, mathematical modeling and comparative analysis methods were applied.

The obtained results demonstrate that the integration of MDP enables more accurate prediction of sequential token selection by taking into account long-term reward functions.

In conclusion, it is justified that such an approach significantly improves the structural quality and contextual relevance of generated text, making it promising for the development of complex Natural Language Processing (NLP) systems.

Keywords: *Language models, Markov Decision Process, MDP, text generation, reinforcement learning, natural language processing, probabilistic models, optimization, machine learning, algorithms.*