KVANT DIQQAT MEXANIZMIGA ASOSLANGAN TRANSFORMER MODELIDA MATNNI NUTQQA AYLANTIRISH TIZIMI

Raximov Nodir Odilovich[1], Primqulov Oybek Dilmurodovich[1], Erkinova Dilnoza Abdurazzoq qizi[1]

[1]Toshkent Axborot Texnalogiyalari Universiteti

K E Y W O R D S

Matnni nutqqa aylantirish, kvant yadro funksiyasi, transformer modeli, kvant diqqat mexanizmi

A B S T R A C T

Ushbu maqolada kvant diqqat mexanizmiga asoslangan Transformer arxitekturasi yordamida matnni nutqqa aylantirish tizimi taklif etiladi. An'anaviy Transformer modellarida qo'llaniladigan diqqat mexanizmi murakkab va chiziqli bo'lmagan bog'lanishlarni ifodalashda imkoniyatlari yetarli darajada emasligi sababli mazkur ishda diqqat mexanizmi kvant yadro funksiyasi bilan takomillashtirildi. Shuningdek, ma'lumotlarni yuqori o'lchamli Hilbert fazosida qayta ishlash orqali nutq signalining fonetik va prosodik xususiyatlarini kengroq modellashtirishga erishiladi. Taqdim etilgan model o'zbek tilining o'ziga xos jihatlari - unli tovushlarning cho'zilishi, bo'g'in ritmi, urg'u va ohangning muhimligi hamda har bir qo'shimcha aloxida ma'no anglatishi hisobga olgan holda ishlab chiqilgan. Modelda vektorlashtirilgan matn kvant diqqat mexanizmi orqali kontekstual bog'lanishlar hosil qilinib, davomiylik, chastota va energiya kabi prosodik parametrlar bashorat qilinadi. Yakuniy natijada mel-spektrogramma generatsiya qilinib vocoder yordamida nutq signaliga aylantiriladi. Olingan natijalar kvant diqqat mexanizmi asoslangan Transformer modeli nutq signalining spektral aniqligi, chastota va energiya barqarorligi hamda tabiiyligini ta'minlashda yuqori aniqlikni ta'minlashini isbotladi.

Kirish

So'nggi yillarda nutq texnologiyalarining rivojlanishi bilan matnni nutqqa aylantirish (Text to Speech, TTS) tizimlarini yanada takomillashtirishga qaratilgan ilmiy tadqiqotlar kengayib bormoqda. Transformer arxitekturasiga asoslangan modellar diqqat mexanizmi yordamida nutq sintezida yuqori natijalarga erishmoqda [1]. Biroq klassik diqqat mexanizmi murakkab va chiziqli bo'lmagan bog'lanishlarni to'liq ifodalashda cheklangan bo'lib, nutq signalining nozik fonetik va prosodik xususiyatlarini chuqur modellashtirishda yetarli darajada samarali emasligini ko'rsatadi. Shu sababli, so'nggi tadqiqotlarda kvant mashinali o'qitish (Quantum Machine Learning) yondashuvlariga katta e'tibor qaratilmoqda [8,9]. Kvant hisoblash modellarida ma'lumotlar yuqori o'lchamli Hilbert fazosida ifodalanadi va bu murakkab bog'lanishlarni aniqlash imkonini sezilarli darajada oshiradi. Ayniqsa, kvant yadro funksiyalaridan foydalanish orqali elementlar orasidagi o'xshashliklarni yanada aniqroq baholash mumkin bo'ladi[8].O'zbek tilida matndan nutq hosil qilish tizimlarini yaratishda o'zbek tilining fonetik va morfologik murakkabliklari muhim rol o'ynaydi. Qo'shimchalarning ketma-ket qo'llanilishi, urg'u va ohangning semantic ma'noga ta'siri hamda bo'g'in tuzilishining o'ziga xosligi TTS modellarida kontekstni yuqori aniqlikda tahlil qilish imkoniyatini beradi.

Mazkur ishda aynan shu yondashuv asosida kvant diqqat mexanizmiga asoslangan Transformer modeli taklif etiladi. Ushbu modelda klassik diqqat mexanizmi kvant yadro funksiyasi bilan boyitilib, kiruvchi ma'lumotlar kvant holatga o'tkaziladi va yuqori o'lchamli fazoda qayta ishlanadi. Natijada model murakkab kontekstual, fonetik va prosodik bog'lanishlarni chuqurroq o'rganish imkoniyatiga ega bo'ladi. Taklif etilgan

(cc) BY

kvant diqqat mexanizmi Transformer arxitekturasi doirasida ishlaydi va nutq sintezi jarayonida kontekstni aniqlash, davomiylikni bashorat qilish hamda akustik parametrlarni shakllantirishda muhim rol o‘ynaydi[1]. Ushbu yondashuv modelning ifodalash qobiliyatini oshirib, generatsiya qilinayotgan nutqning tabiiyligi va aniqligini yaxshilashga xizmat qiladi. Shunday qilib, ushbu maqolaning asosiy maqsadi kvant diqqat mexanizmiga asoslangan Transformer modelini ishlab chiqish va uning matnni nutqqa aylantirish tizimlarida qo‘llanish samaradorligini tadqiq etishdan iborat.

Quyida kvant diqqat mexanizmiga asoslangan Transformer modeli matnni nutqqa aylantirish jaroyoni ko’rib chiqilgan:

Kvant diqqat mexanizmiga asoslangan Transformer TTS

1-bosqich. Matnni vektorlashtirish va pozitsion kodlash. Modelga kiritiladigan matn vektor ko‘rinishiga o‘tkaziladi va tokenlar haqida ma’lumot qo’shiladi. Kirish matni quyidagicha tokenlar ketma-ketligi orqali ifodalanadi.

$$x = (x_1, \dots, x_T) \qquad (1)$$

Vektorlashtirish jaroyonida har bir token vektor ko’rinishiga o‘tkaziladi va quyidagicha ifodalanadi:

$$e_t = E[x_t] \qquad (2)$$

Transformer modelida tokenlar ketma-ketligini saqlash uchun pozitsion kodlash qo’shiladi:

$$\tilde{z}_t = e_t + PE(t) \qquad (3)$$

2-bosqich. Matnni kontekstual kodlash (Self-attention). Matn ketma-ketligidagi tokenlar qanday bog‘langanini o‘rganadi va kontekstual xususiyatlarga ajratiladi [1]. Self-attention qatlamida uzatiladigan vektorlar quyidagicha ifodalanadi:

$$Q = ZW_Q, \quad K = ZW_K, \quad V = ZW_V \qquad (4)$$

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (5)$$

$$MultiHead(Q,K,V) = concat(head_1, \dots, head_h)W_O \qquad (6)$$

$$Z' = LayerNorm(Z + MultiHead(Q,K,V)) \qquad (7)$$

$$H = LayerNorm(Z' + FFN(Z')) \qquad (8)$$

3-bosqich. Kvant yadro mexanizmi. Ushbu bosqichda encoder yuqori o’lchamli bog‘lanishlarni o‘rganish uchun kvant yadro holatda kodlanadi va quyidagicha ifodalanadi:

$$H_i \rightarrow |\phi(H_i)\rangle \qquad (9)$$

Kodlangan encoder orqali kvant yadro funksiyasi hisoblanadi:

$$K(H_i, H_j) = \langle \phi(H_i) | \phi(H_j) \rangle^2 \qquad (10)$$

$$a_{ij} = \frac{K(H_i,H_j)}{\sum_{k=1}^{T} K(H_i,H_k)} \qquad (11)$$

Hosil bo’lgan natijaviy kvant mexanizmi shunday hisoblanadi:

$$H_Q = \sum_j a_{ij} V_j \qquad (12)$$

4-bosqich. Davomiylikni bashorat qiluvchi model. Har bir matn davomiyligi va uzunligi kvant regressiya yondashuvi yordamida aniqlanadi.

$$\hat{d} = f_{dur}^{Q}(H_Q) \qquad (13)$$

bunda $\hat{d}$- bashorat qilingan davomiylik, H_Q- kvant attention mexanizmi orqali hosil bo’lgan reprezentatsiya.

5-bosqich. Kvant tovush amplitudasi va energiyasini bashorat qilish. (Kvant pitch va energy predictor). Nutqning tovush amplitudasi va energiyasi kvant hisoblash sxemasi yordamida bashorat qilinadi:

$$\hat{p} = f_{pitch}^{Q}(H'), \quad \hat{e} = f_{energy}^{Q}(H') \qquad (14)$$

Embedding funksiyasi modelni vektor ko’rinishiga o‘tkazadi va har bir vektorga yangi ma’lumot qo’shib boradi:

$$H'' = H' + Emb_p(\hat{p}) + Emb_e(\hat{e}) \qquad (15)$$

6-bosqich. Dekodlash. (Akustik mel-spektograma hosil qilish). Mazkur bosqichda hosil

(cc) BY

qilingan vektor ko‘rinishidagi ifoda mel-spektogrammaga o’tkaziladi [2,6]:

$$M = Decoder(H'') \quad (16)$$

7-bosqich. Vocoder (Mel-spektogramma kadrlarini audioga o‘zgartirish). Hosil qilingan mel-spektrogramma kadrlari vocoder yordamida audio signalga aylantiriladi:

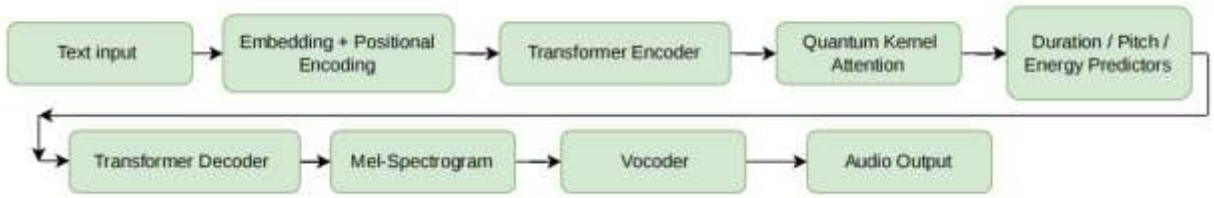
$$y = Vocoder(M) \quad (17)$$

8-bosqich. Yo’qotish funksiyasini hisoblash. So‘nggi bosqichda tayyor bo’lgan audio signal yo’qotish funksiyalari orqali tozalanadi aniqlikka erishiladi:

$$\mathcal{L}_{mel} = \|M - \widehat{M}\|_1 \quad (18)$$
$$\mathcal{L}_{dur} = \|d - \hat{d}\|_2^2 \quad (19)$$
$$\mathcal{L}_{pitch} = \|p - \hat{p}\|_2^2 \quad (20)$$
$$\mathcal{L}_{energy} = \|e - \hat{e}\|_2^2 \quad (21)$$
$$\mathcal{L} = \mathcal{L}_{mel} + \lambda_1\mathcal{L}_{dur} + \lambda_2\mathcal{L}_{pitch} + \lambda_3\mathcal{L}_{energy} \quad (22)$$

Kvant transformer tizimining umumiy ishlash bosqichlari quyidagicha izohlanadi

$$Text \rightarrow Emb + PE \rightarrow Encoder \rightarrow Quantum\ Attention \rightarrow Quantum\ Duration\ /Pitch\ /Energy \rightarrow Decoder \rightarrow Mel \rightarrow Vocoder \rightarrow Audio$$



4-rasm. Kvant diqqat mexanizmiga asoslangan Transformer modeli

Tadqiqot natijalari

Ushbu modellar 6784 ta jami 78 soatlik o’zbek tiliga xos bo’lgan gaplardan tashkil topgan ma’lumotlar to’plamidan foydalanib sinovdan o’tkazildi. Ma’lumotlar to’plamini 80% training va 20% test uchun tanlab olindi. Sinov natijasida modellar quyidagi baholash metrikalari asosida hisoblandi.

Baholash usullari quyidagicha aniqlanadi:

1.Chastota uchun o’rtacha kvadratik hatolik hisoblash quyidagicha aniqlanadi:

$$RMSE_{F0} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(F0_t - \widehat{F0}_t)^2} \quad (23)$$

bunda $F0_t$-haqiqiy chastota, $\widehat{F0}_t$- bashorat qilingan chastota, T-freym soni.

2. Amplituda o’rtacha kvadratik hatolikni hisoblash uchun avvalo amplituda hisoblanadi:

$$E_t = \sqrt{\frac{1}{N}\sum x^2} \quad (24)$$

$$RMSE_E = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(E_t - \hat{E}_t)^2} \quad (25)$$

3. Haqiqiy chastota $F0_t$ ni bashorat qilingan chastota $\widehat{F0}_t$ ga bog’liqligi korelatsiya koeffisenti orqali aniqlanadi:

$$Corr = \frac{\sum(F0_t - \overline{F0})(\widehat{F0}_t - \overline{F0})}{\sqrt{\sum(F0_t - \overline{F0})^2}\sqrt{\sum(\widehat{F0}_t - \overline{F0})^2}} \quad (26)$$

4. MCD (Mel-Cepstral Distortion) haqiqiy va bashorat qilingan mel-spektral koeffisent o‘rtasidagi farqni o‘lchaydi:

$$MCD = \frac{10}{ln\,10}\sqrt{2\sum_{d=1}^{D}(c_d - \hat{c}_d)^2} \quad (27)$$

bunda c_d-haqiqiy mel-cepstral koeffitsient, $\hat{c}_d$-model bashorat qilgan koeffitsient D-koeffitsientlar soni.

5. LSD (Log-Spectral Distance) log spektr orasidagi farqlarni tahlil qilish quyidagicha aniqlanadi:

$$LSD = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(S(k) - \hat{S}(k))^2} \quad (28)$$

(cc) BY

bunda S(k)- haqiqiy spektr , $\hat{S}(k)$-model bashorat qilgan spektr, K –chastota binlari soni.

6. MOS (Mean Opinion Score) shaxsga bo'gliq bo'lgan baholash ko'rsatkichi hisoblanib, tinglovchi tomonidan baholanadi [11].

$$MOS = \frac{1}{N}\sum_{i=1}^{N} s_i \qquad (29)$$

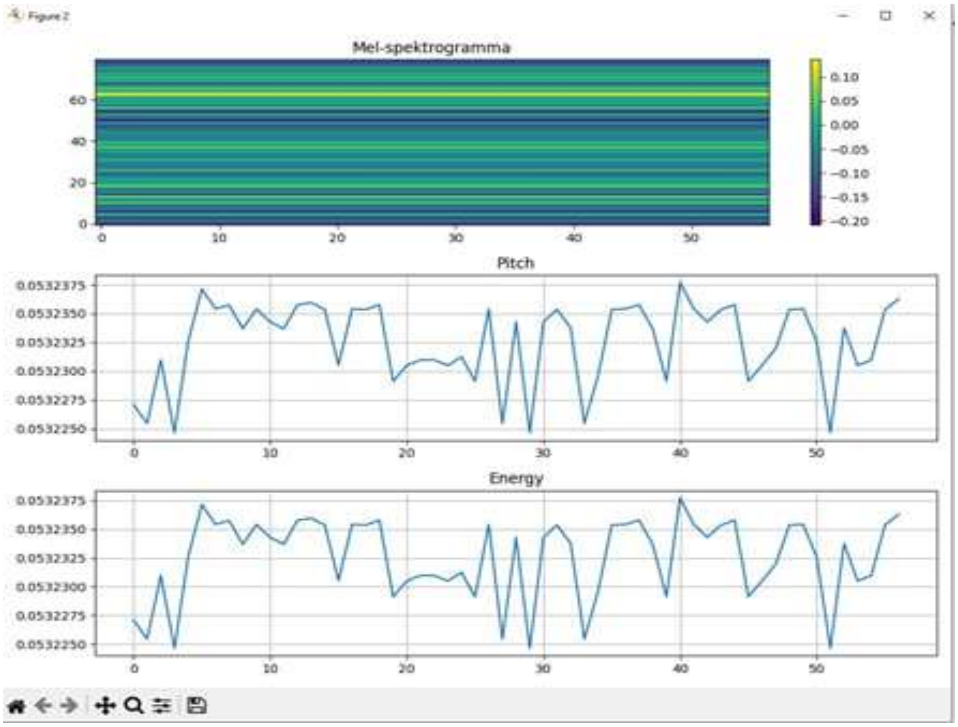
bunda s_i-foydalanuvchi bahosi, N- baholovchilar soni.

7. PESQ (Perceptual Evaluation of Speech Quality) - xalqaro stardantga asoslangan baholash usuli hisoblanib, nutq sifati ko'rsatkichini aniqlash uchun ishlatiladi:

$$PESQ = a_0 + a_1 D_{sym} + a_2 D_{asym} \quad (30)$$

bunda D_{sym} - simmetrik buzilish, D_{asym} - assimetrik buzilish, a_0, a_1, a_2 - model konstantalari.

Quyida TTS tizimining Kvant diqqat mexanizmiga asoslangan Tranformer modellari hosil qilgan mel-spektogramma amplituda va chastota grafigi keltirilgan.



7-rasm. Kvant diqqat mexanizmiga asoslangan Transformel modeli asosida TTS modeli

Grafik natijalardan ko'rinadiki, kvant modelda mel-spektrogramma silliq, pitch va energy esa barqaror o'zgarishga ega. Bu nutq signalining tabiiyligi va prosodik aniqligi oshganligini ko'rsatadi.

1-jadval.

Kvant diqqat mexanizmiga asoslangan Transformer modelining baholash natijalari

Model	F0 RMSE	Energy RMSE	MCD	LSD	MOS	PESQ
Q-Transformer	8.12	0.097	4.36	1.08	4.67	4.15

Jadval natijalari shuni ko'rsatadiki, kvant diqqat mexanizmiga asoslangan Transformer modeli nutq signalining asosiy akustik va prosodik ko'rsatkichlari bo'yicha yuqori aniqlikni ta'minlaydi. F0 RMSE va Energy RMSE qiymatlarining pastligi chastota va energiya

(cc) BY

bashoratining aniqligini bildiradi hamda MCD va LSD ko‘rsatkichlarining pastligi spektral farqning kamayganini, ya’ni generatsiya qilingan nutqning tabiiyligi oshganini ko‘rsatadi. MOS va PESQ qiymatlarining yuqoriligi esa model tomonidan hosil qilingan nutq signalining eshitish sifatini yuqori darajada ekanligini tasdiqlaydi.

Xulosa

Ushbu maqolada kvant diqqat mexanizmiga asoslangan Transformer modeli yordamida matnni nutqqa aylantirish tizimi taklif etildi. Taklif etilgan yondashuvda klassik diqqat mexanizmi kvant yadro funksiyasi orqali yoritilib, ma’lumotlar yuqori o‘lchamli fazoda qayta ishlanishi ta’minlandi. Bu esa nutq signalining murakkab fonetik va prosodik xususiyatlarini yanada aniq modellashtirish imkonini berdi. Tadqiqot natijalari kvant diqqat mexanizmi asosidagi model nutq signalining chastota, energiya va spektral ko‘rsatkichlari bo‘yicha yuqori aniqlikka erishganini ko‘rsatdi. Ko‘rsatilgan F0 RMSE, Energy RMSE, MCD va LSD qiymatlarining pastligi generatsiya qilingan nutq signalining akustik jihatdan aniq va tabiiy ekanligini bildiradi. MOS va PESQ baholarining yuqoriligi esa model tomonidan hosil qilingan nutqning eshitish sifatini tasdiqlaydi. Kvant yadro funksiyasidan foydalanish modelning murakkab chiziqli bo’lmagan bog‘lanishlarni aniqlash qobiliyatini oshirib, nutq sintezida prosodik barqarorlikni ta’minladi.

Xulosa qilib aytganda, kvant diqqat mexanizmiga asoslangan Transformer modeli matnni nutqqa aylantirish tizimlarida yuqori sifatli va tabiiy nutq generatsiyasini ta’minlovchi samarali yondashuv ekanligi isbotlandi. Kelgusida ushbu modelni real kvant hisoblash tizimlarida sinovdan o‘tkazish hamda gibrid kvant-klassik arxitekturalarni rivojlantirish muhim ilmiy yo‘nalishlardan biri hisoblanadi.

Foydalanilgan adabiyotlar.

1. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention Is All You Need // *Advances in Neural Information Processing Systems*, 2017.
2. Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang. Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions // *ICASSP*, 2018.
3. Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, Bryan Catanzaro. FastSpeech: Fast, Robust and Controllable Text to Speech // *NeurIPS*, 2019.
4. Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, Bryan Catanzaro. DiffWave: A Versatile Diffusion Model for Audio Synthesis // *ICLR*, 2021.
5. Ho, J., Jain, A., Abbeel, P. Denoising Diffusion Probabilistic Models // *NeurIPS*, 2020.
6. Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves. Generative Model for Raw Audio // *arXiv preprint arXiv:1609.03499*, 2016.
7. Tokuda, K., Zen, H., Black, A.W. An HMM-based Speech Synthesis System Applied to English // *IEEE Speech Synthesis Workshop*, 2002.
8. Schuld, M., Sinayskiy, I., Petruccione, F. An Introduction to Quantum Machine Learning // *Contemporary Physics*, 2015.
9. Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, Seth Lloyd. Quantum Machine Learning // *Nature*, 2017.
10. Rabiner, L., Juang, B.H. Fundamentals of Speech Recognition. Prentice Hall, 1993.
11. Jurafsky, D., Martin, J.H. Speech and Language Processing. Pearson, 2020.

(cc) BY