

Застосування нейронних мереж для вирішення завдання генерації музики

Дорогий Я. Ю., к.т.н., доц., ORCID [0000-0003-3848-9852](https://orcid.org/0000-0003-3848-9852)

e-mail argusyk@gmail.com

Цуркан В. В., к.т.н., ORCID [0000-0003-1352-042X](https://orcid.org/0000-0003-1352-042X)

e-mail v.v.tsurkan@gmail.com

Бельдяга К. Д., ORCID [0000-0002-6456-6981](https://orcid.org/0000-0002-6456-6981)

e-mail kirill.beldyaga@gmail.com

Національний технічний університет України

"Київський політехнічний інститут імені Ігоря Сікорського" kpi.ua

Київ, Україна

Реферат—В статті розглядається можливість використання нейронних мереж як основи генератора музики. Під генерацією музики розуміють процес створення послідовностей музичних нот, що відносить дане завдання до задач моделювання послідовностей, як і моделювання мови. Актуальність розвитку даної теми пов'язана з необхідністю написання нової музики та знаходження нових мотивів, що спричинено виробництвом великої кількості фільмів та потребами людей. Основною вимогою до системи була генерація поліфонічної музики, що є послідовністю кортежів нот, тобто багаторозмірним об'єктом, що потребує використання породжувальних моделей, які здатні обчислювати ймовірнісний розподіл по відношенню до рекурентних моделей, які, в свою чергу, дають змогу відтворити довгострокові залежності в послідовності багаторозмірних об'єктів. Для цього в статті наведено необхідні теоретичні відомості. Розглянуто існуючі рішення і визначено їх переваги та недоліки. Проаналізовано та представлено варіанти можливих топологій, алгоритми їх тренування та генерації. Описано результати експериментальних досліджень, на підставі яких виконано порівняння та визначено переваги та недоліки кожної з них.

Бібл. 11, рис. 8, табл. 1.

Ключові слова — нейронні мережі; музика; моделювання послідовностей; генерація музики; машинне навчання.

I. ВСТУП

Генератор музики може бути використаний будь-де в індустрії медіарозваг – від інструменту, що допомагає музикантам, до повноцінної системи, яка здатна в реальному часі в залежності від дій в фільмі генерувати музичний фон.

II. АНАЛІЗ ЛІТЕРАТУРНИХ ДАНИХ І ПОСТАНОВКА ПРОБЛЕМИ

Моделювання музики лише починає розвиток, на відміну від завдань розпізнавання та відтворення мови, тому повноцінних рішень, які б задовольняли потреби, наприклад, композиторів, що шукають джерело натхнення чи мотивацію, поки що не існує. Проте є достатньо досліджень та експериментальних рішень, зокрема, компанія Google представила рішення Magenta [1] та WaveNet [2]. Перше використовується для генерації монофонічної музики, а будова іншого дещо складніша. WaveNet базується на згорткових нейронних мережах і дозволяє генерувати музику в форматі "raw" та використовувати звуки різних інструментів одночасно, що є надзвичайно складним. Що і відображається на економічних витратах – для генерації однієї секунди звуку необхідний час тренування складає хвилину. Також слід звернути увагу на експериментальне рішення [3],

в якому основна увага приділяється проблемі інваріантності за тональністю, оскільки всі рекурентні мережі є інваріантні за часом, але не за тональністю. З цієї причини в ньому використовується принцип згорткових нейронних мереж, а у якості згорткового ядра використовується двонаправлена довготривала короткочасна пам'ять (ДКЧП), яка є поєднанням двох стеків ДКЧП. Перший складається з ДКЧП, що мають рекурентні зв'язки по вісі часу, та призначений для того, щоб система була інваріантна за часом, і визначає часовий проміжок ноти, а інший, елементи якого мають рекурентні зв'язки по вісі тональності, визначає саму ноту. Результатом такої складної системи є вектор ймовірностей, що ті ноти будуть згенеровані.

Рекурентна нейронна мережа (РНМ) – це штучна нейронна мережа, у якій вузли використовують вихідні сигнали у якості вхідних, утворюючи цикл, розглядаючи який в часі, можна отримати ланцюг елементів, кожен з яких відповідає часовому кроку, виходом нейрона на кожному з яких є значення функції гіперболічного тангенса, аргументом якої є сума зважених матрицями вагових коефіцієнтів векторів значень входу x_t , та виходу на минулому кроці y_{t-1} :

$$y_t = \tanh(b_y + W_{xy}x_t + W_{yy}y_{t-1}),$$

де b_y – вектор зсувів; W_{xy} , W_{yy} – матриці вагових коефіцієнтів [4].



Однак РНМ мають проблему зникання градієнту. При використанні методу зворотного поширення в часі градієнт (помилка) експоненційно спадає. Цю проблему описав Джозеф Хокрейтер у 1991 році [5], представивши в 1997 [6] вирішення даної проблеми, яке має назву «довга короткочасна пам'ять».

Довга короткочасна пам'ять – особливий вид РНМ, яка пристосована для вирішення проблеми довготривалих залежностей. Як і РНМ, ДКЧП може бути представлена у якості ланцюга, проте структура елементів складніша. Елемент ДКЧП складається із комірки стану c_t , воріт f_t , що регулюють дані, що необхідно видалити, вхідних воріт i_t , що фільтрують вхідні значення x_t для зберігання в комірці, а також вихідних o_t , які регулюють інформацію, яку необхідно отримати в результаті [7].

$$\begin{aligned} f_t &= \sigma(b_f + W_{cf}c_{t-1} + W_{yf}y_{t-1} + W_{xf}x_t), \\ i_t &= \sigma(b_i + W_{ci}c_{t-1} + W_{yi}y_{t-1} + W_{xi}x_t), \\ c_t &= f_t c_{t-1} + i_t \tanh(b_c + W_{cx}x_t + W_{cy}y_{t-1}), \\ o_t &= \sigma(b_o + W_{co}c_t + W_{yo}y_{t-1} + W_{xo}x_t), \\ y_t &= o_t \tanh(c_t). \end{aligned} \quad (1)$$

Зазвичай для ефективного моделювання послідовностей використовується РНМ або її різновиди, зокрема, ДКЧП, оскільки вони дають змогу відтворити довгострокові залежності, тобто всю логічну послідовність, що є обов'язковим в генерації музики, наприклад: деякі ноти на початку композиції дуже часто впливають на останні або коли якась музична фраза повторюється декілька разів та є обов'язковою для донесення почуттів та емоцій композитора слухачеві.

Для генерації монофонічної музики достатньо використати РНМ або ДКЧП-структури, проте для поліфонічної просте прогнозування наступної ноти недостатнє, оскільки записується вона у вигляді кортежів музичних нот, що є багатовимірним об'єктом. Тому нейронна мережа повинна прогнозувати умовний розподіл наступного кроку, використовуючи попередній. Для генерації поліфонічної музики це необхідно, оскільки можна визначити ймовірність того, що певна нота з'явиться з деякими іншими в один і той же час. Використання лише РНМ в такому випадку є неекономічним.

Вирішенням даної проблеми є поєднання РНМ з породжувальними моделями, які дозволяють визначити спільний розподіл ймовірності над послідовністю. В якості останньої досить часто використовується обмежена машина Больцмана (ОМБ). А для збільшення варіаційної нижньої межі функції правдоподібності можна ієрархічно представляти дані, що досягається поєднанням мереж, що мають можливість визначати довготривалі залежності, наприклад, РНМ або ДКЧП, з глибинною мережею переконань (ГМП).

Обмежена машина Больцмана (ОМБ) – породжувальна модель, яка визначає ймовірнісний розподіл, через функцію енергії системи при її конфігурації параметрів [8].

$$p(v, h) = \frac{e^{-E(v, h)}}{Z} = \frac{e^{-E(v, h)}}{\sum_v e^{-E(v, h)}} \quad (2)$$

де v, h – вектори вхідного (видимого) та прихованого шарів; Z – нормалізуючий фактор; $E(v, h)$ – енергетична функція, що визначається наступною формулою:

$$E(v, h) = - \sum_i b_i v_i - \sum_j c_j h_j - \sum_i \sum_j w_{ij} v_i h_j,$$

де b, c – вектори зсувів видимого та прихованого шару, а W – матриця вагових коефіцієнтів.

З вищезгаданих формул випливає, що чим нижче енергія конфігурації системи, тим вище ймовірність переходу її у цей стан.

Маргінальна ймовірність того, що v відноситься до $F(v)$.

$$p(v) = \frac{e^{-F(v)}}{Z},$$

де $F(v)$ – деяка функція вільної енергії, що є енергією по всім конфігураціям прихованого шару та визначається виразом (3).

$$F(v) = b_v^T v - \sum_i \log(1 + e^{b_i + W_{iv} v_i}). \quad (3)$$

Глибинна мережа переконань (ГМП) – це породжувальна модель, що складається з декількох прихованих шарів. ГМП може бути композицією таких моделей, як ОМБ або автокодувачів [9]. В цій роботі буде використано ОМБ. В тому випадку, як ГМП формується з двох або більше з'єднаних ОМБ, необхідно до прихованого шару додати ще один прихований шар, в результаті чого послідовне з'єднання другого прихованого, першого прихованого та видимого шарів утворює дві ОМБ таким чином, що для верхньої ОМБ видимим шаром є прихований шар нижньої ОМБ.

Спільний розподіл ймовірностей в цій моделі обчислюється наступною формулою:

$$P(v, h^{(1)}, \dots, h^{(n)}) = \left(\prod_{k=0}^{n-2} P(h^{(k)} | h^{(k+1)}) \right) P(h^{(n-1)}, h^{(n)}),$$

де n – кількість прихованих шарів.

III. МЕТА І ЗАВДАННЯ ДОСЛІДЖЕННЯ

Метою дослідження є порівняння декількох топологій нейронних мереж, які дозволяють генерувати послідовності музичних нот, та вибір однієї з них, що буде використана як основа генератора музики.

Дослідження спрямоване на розв'язання наступних завдань:

- визначити варіанти можливих топологій, використовуючи наведені теоретичні відомості,
- навести алгоритми тренування та генерації для всіх варіантів структур,
- порівняти топології, провівши експериментальні дослідження,
- дотримуватися наступних вимог до генератора: музика має бути поліфонічною та мати якомога менше випадкових нот, тобто помилкових зв'язків між ними, що буде обрховано за допомогою перехресної ентропії.



IV. ОСНОВНА ЧАСТИНА

А. Синтез топології

Оскільки метою цієї роботи є порівняння декількох топологій, для виконання даної роботи було взято чотири архітектури – РНМ-ОМБ [10], ДКЧП-ОМБ, РНМ-ГМП та ДКЧП-ГМП.

РНМ-ОМБ – це послідовність з поєднань ОМБ та одного вузла РНМ. На кожному часовому кроці ОМБ виконує ймовірнісний розподіл, а РНМ формує поточний вихід, сумуючи зважені значення виходу поточного кроку та входу поточного кроку, що є видимим шаром ОМБ. Прихований шар ОМБ складається з 150 нейронів, а видимий – з 88, що відповідає кількості клавіш на клавіатурі піаніно. В цьому випадку використовується бінарна ОМБ, тобто $y_t \in \{0,1\}$. Таким чином, одиниця означає, що клавіша має бути натиснута, а нуль – піднята. Структура такої нейронної мережі представлена на рис. 1.

Параметри ОМБ визначаються наступними формулами:

$$b_{h_t} = b_h + W_{yh} y_{t-1}, \quad (4)$$

$$b_{v_t} = b_v + W_{yv} y_{t-1}, \quad (5)$$

де $b_h, b_v, b_h^{(t)}, b_v^{(t)}$ – вектори зсувів видимого та прихованих шарів попереднього та поточного кроку; W_{yh}, W_{yv} – матриці вагових коефіцієнтів; y_{t-1} – значення РНМ на попередньому кроці.

Значення виходу вузла формується наступним виразом:

$$y_t = \tanh(W_{vx} v_t + W_{yy} y_{t-1} + b_y), \quad (6)$$

де W_{vy}, W_{yy} – матриці вагових коефіцієнтів між видимим шаром ОМБ і прихованим РНМ та між вузлами РНМ на попередньому і поточному кроках, відповідно; v_t – вектор значень видимого шару ОМБ; b_y – вектор зсувів прихованого шару РНМ.

На протизагу топології РНМ-ОМБ було реалізовано архітектуру ДКЧП-ОМБ, принцип роботи якої

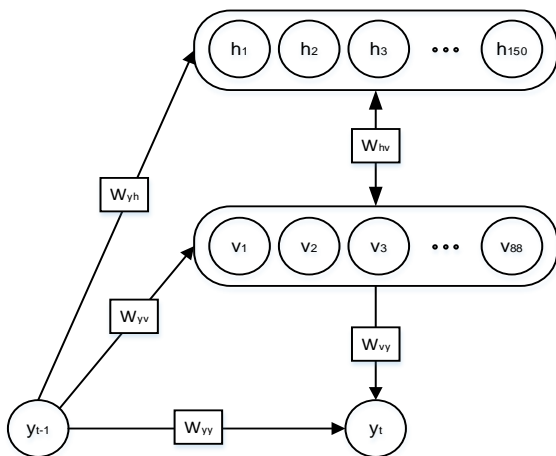


Рис. 1 Структурна схема топології РНМ-ОМБ

аналогічний до першої, однак є певні відмінності в обрахунку поточного значення вузла РНМ. Параметри ОМБ обраховуються виразами (11) та (5). Поточний вхід визначається виразом (7), що передається аргументом до виразів (2-5).

$$x_t = \sigma(b_x + W_{vx} v_t + W_{yx} y_{t-1}), \quad (7)$$

де b_x – вектор зсувів входів вузла ДКЧП; W_{vx} – матриця вагових коефіцієнтів між шарами видимого ОМБ та прихованого ОМБ; W_{yx} – матриця ваг між попереднім та поточним вузлами ДКЧП; y_{t-1} – вихід ДКЧП на попередньому кроці. Поточний вхід обраховується виразом (1). Структура описаної вище нейронної мережі наведена на рис. 2.

Для глибокого вивчення структури мелодії було використано глибинну мережу переконань (ГМП) в поєднанні з РНМ. Параметри ГМП визначаються наступними виразами:

$$b_{h_t}^{(n)} = b_h^{(n)} + W_{yh^{(n)}} y_{t-1}, \quad (8)$$

$$b_{v_t} = b_v + W_{yv} y_{t-1},$$

де n – кількість прихованих шарів ГМП; $b_{h_t}^{(n)}, b_h^{(n)}$ – вектори зсувів прихованих шарів в поточному та попередньому кроках; b_{v_t}, b_v – вектори зсувів видимого шару в моменті часу t і $t-1$; $W_{yh^{(n)}}$ – матриця вагових коефіцієнтів між попереднім вузлом РНМ та поточним прихованим шаром n ГМП.

Оскільки ГМП складається із двох ОМБ, вираз (8) можна представити у вигляді наступних:

$$b_{h_t}^{(2)} = b_h^{(2)} + W_{yh^{(2)}} y_{t-1},$$

$$b_{h_t}^{(1)} = b_h^{(1)} + W_{yh^{(1)}} y_{t-1}, \quad (18)$$

значення виходу формується виразом (6). Структура РНМ-ГМП представлена на рис. 3.

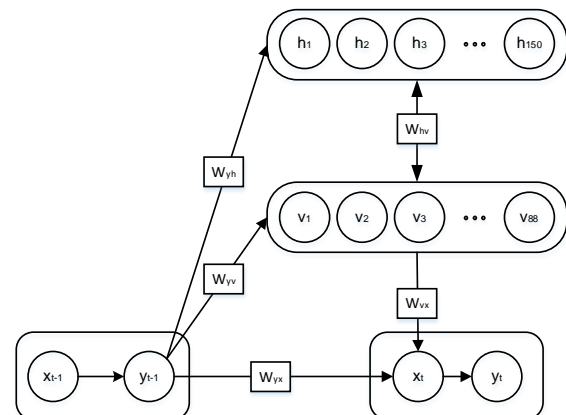


Рис. 2 Структурна схема топології ДКЧП-ОМБ

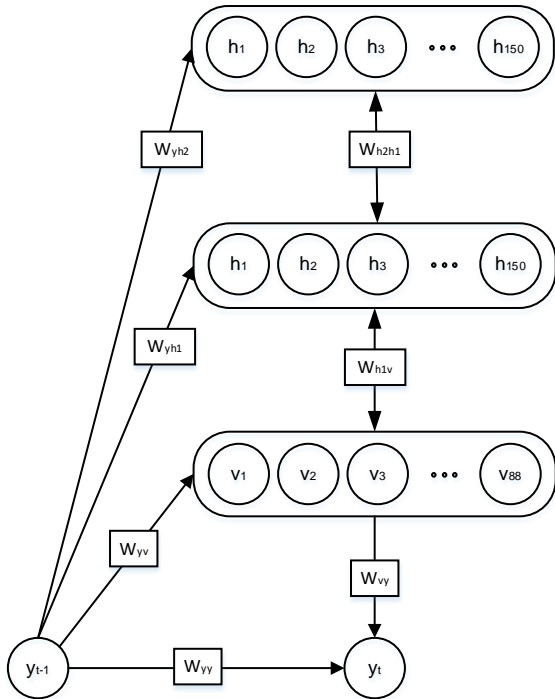


Рис. 3 Структурна схема топології РНМ-ГМП

Аналогічно до ДКЧП-ОМБ, на протипагу РНМ-ГМП було обрано ДКЧП-ГМП. Відмінність від РНМ-ГМП лише в тому, що вхід ДКЧП формується виразом (7), а вихід – виразом (1). Структура такої мережі представлена на рис. 4.

В. Тренування

В структурах всіх топологій є ОМБ, при тренуванні якої застосовується метод порівняльної розбіжності:

$$\frac{\partial(-\log p(v_t))}{\partial\theta} = \frac{\partial F(v_t)}{\partial\theta} - \frac{\partial F(\tilde{v}_t)}{\partial\theta}, \tag{9}$$

де $-\log p(v_t)$ – негативна функція правдоподібності; v_t – тренувальний зразок, \tilde{v}_t – згенерований зразок; $\theta = \{c, b, W\}$ – параметри системи. Даний метод дозволяє обрахувати, на скільки згенерований зразок є схожим на тренувальний.

Використовуючи властивість ОМБ обраховувати ймовірнісний розподіл (2), можна обчислити умовний розподіл, коли $h=1$ та $v=1$:

$$p(h=1|v) = \sigma(w_v v + b_j), \tag{10}$$

$$p(v=1|h) = \sigma(w_h h + c_i). \tag{11}$$

З чого випливає, що значення прихованого та видимого шарів подібне до значень умовних розподілів:

$$h_{t+1} \sim \sigma(w_{vh} v_t + b), \tag{12}$$

$$v_{t+1} \sim \sigma(w_{hv} h_{t+1} + c). \tag{13}$$

Таким чином можна обчислити \tilde{v}_t . Вищеописаний процес має назву «відбір проб за Гібсом».

Алгоритм тренування для РНМ-ОМБ.

- 1) Ініціалізувати змінні, які є параметрами системи.

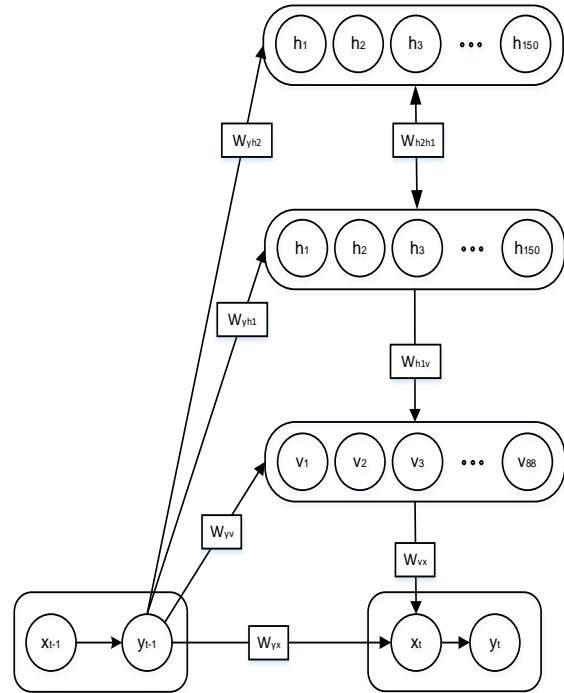


Рис. 4 Структурна схема топології ДКЧП-ГМП

- 2) Визначити конфігурацію системи на основі тренувальних даних на першому часовому кроці.
- 3) Використовуючи формули (4) та (5), визначити параметри ОМБ на поточному кроці.
- 4) Виконати ітерацію відбору проб «за Гібсом», використовуючи формули (10-11) та (12-13), щоб згенерувати видимий шар v_t ОМБ.
- 5) Порівняти згенерований вектор нот з наявними в мелодії, застосувавши метод порівняльної розбіжності (9).
- 6) На основі обрахованого значення функції витрат методом зворотного поширення обрахувати нові параметри системи.
- 7) За допомогою виразу (6) обчислити значення РНМ, використовуючи оновлені параметри ОМБ.
- 8) Повторювати кроки 3-7 до закінчення тренування.

Алгоритм тренування для ДКЧП-ОМБ.

- 1) Ініціалізувати змінні, які є параметрами системи.
- 2) Визначити конфігурацію системи на основі тренувальних даних на першому часовому кроці.
- 3) Використовуючи формули (4) та (5), визначити параметри ОМБ на поточному кроці.



- 4) Виконати ітерацію відбору проб «за Гібсом», використовуючи формули (10-11) та (12-13), щоб згенерувати видимий шар v_t ОМБ.
- 5) Порівняти згенерований вектор нот з наявними в мелодії, застосувавши метод порівняльної розбіжності (9).
- 6) На основі обрахованого значення функції витрат методом зворотного поширення обрахувати нові параметри системи.
- 7) За допомогою виразів (7) та (2-6) обчислити значення ДКЧП, використовуючи оновлені параметри ОМБ.
- 8) Повторювати кроки 3-7 до закінчення тренування.

Алгоритм тренування для РНМ-ГМП.

- 1) Ініціалізувати змінні, які є параметрами системи.
- 2) Визначити конфігурацію системи на основі тренувальних даних на першому часовому кроці.
- 3) Використовуючи формули (4) та (5), визначити параметри ОМБ на поточному кроці.
- 4) Виконати ітерацію відбору проб «за Гібсом», використовуючи формули (10-11) та (12-13), щоб згенерувати видимий шар v_t ОМБ. Для верхньої ОМБ видимим шаром є прихований нижчої ОМБ.
- 5) Порівняти згенерований вектор нот з наявними в мелодії, застосувавши метод порівняльної розбіжності (9).
- 6) Повторити дії 4 та 5 для кожного з прихованих шарів.
- 7) На основі обрахованого значення функції витрат методом зворотного поширення обрахувати нові параметри системи.
- 8) За допомогою виразів (6) обчислити значення ДКЧП, використовуючи оновлені параметри ОМБ.
- 9) Повторювати кроки 3-8 до закінчення тренування.

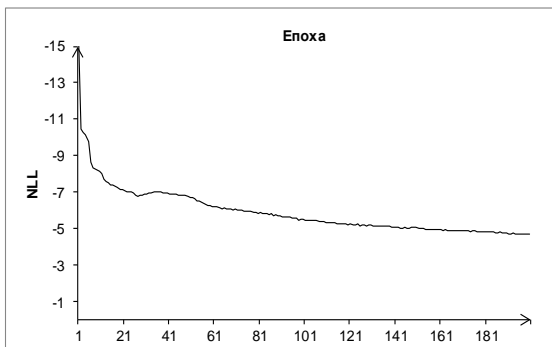


Рис. 5 Графік падіння цільової функції РНМ-ОМБ

Алгоритм тренування для ДКЧП-ГМП.

- 1) Ініціалізувати змінні, які є параметрами системи.
- 2) Визначити конфігурацію системи на основі тренувальних даних на першому часовому кроці.
- 3) Використовуючи формули (4) та (5), визначити параметри ОМБ на поточному кроці.
- 4) Виконати ітерацію відбору проб «за Гібсом», використовуючи формули (10-11) та (12-13), щоб згенерувати видимий шар v_t ОМБ. Для верхньої ОМБ видимим шаром є прихований нижчої ОМБ.
- 5) Порівняти згенерований вектор нот з наявними в мелодії, застосувавши метод порівняльної розбіжності (9).
- 6) Повторити дії 4 та 5 для кожного з прихованих шарів.
- 7) На основі обрахованого значення функції витрат методом зворотного поширення обрахувати нові параметри системи.
- 8) За допомогою виразів (7) та (2-6) обчислити значення ДКЧП, використовуючи оновлені параметри ОМБ.
- 9) Повторювати кроки 3-8 до закінчення тренування.

С. Генерація

Алгоритм генерації:

- 1) Ініціалізувати необхідні змінні. Видимі рівні ОМБ заповнити нулями або видимим шаром v_{t-1} , взятим на попередньому кроці тренування.
- 2) Визначити v_t , виконати k кроків відбору проб «за Гібсом».
- 3) Обчислити значення виходу рекурентної частини, для РНМ за виразом (6), для ДКЧП – за (7).

Для всіх структур алгоритм однаковий.

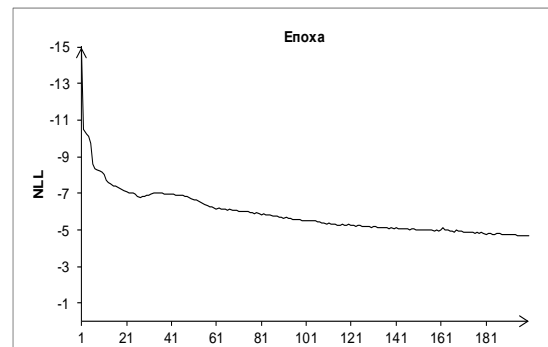


Рис. 6 Графік падіння цільової функції ДКЧП-ОМБ

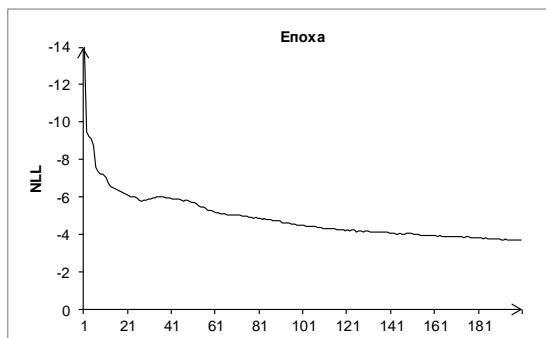


Рис. 7 Графік падіння цільової функції RHM-ГМП

D. Оцінка якості моделей

Ціль генератора – спрогнозувати розподіл ймовірностей, що певні ноти мають з'явитися, а які – ні, а такий метод для обчислення помилки, як середньої квадратичної помилки, не підходить, тому в даному випадку слід застосувати вартість перехресної ентропії:

$$CEC = -\frac{1}{N} \sum_{n=1}^N x_n \ln \hat{x}_n + (1-x_n) \ln(1-\hat{x}_n),$$

де N – кількість елементів вектору v ; x_n – ймовірність того, що тренувальний зразок рівний 1, а $1-x_n = 0$; \hat{x}_n – ймовірність того, що згенерований зразок рівний 1, а $1-\hat{x}_n = 0$.

Функцією втрат є негативна функція правдоподібності, яка оптимізується методом стохастичного градієнтного спуску, з використанням порівняльної розбіжності:

$$NLL = -\log p(v_i).$$

Точність показує, скільки було згенеровано правильних результатів та визначається виразом:

$$ACC = \frac{P}{Q},$$

де P – кількість правильно згенерованих значень, а Q – загальна кількість згенерованих значень.

Важливим критерієм вибору основної топології для генератора музики є час тренування. Спираючись на велику кількість дослідів, можна порахувати середній час, що був необхідний на тренування різних топологій.

V. ЕКСПЕРИМЕНТИ ТА РЕЗУЛЬТАТИ

В якості тренувальних даних використовувалась «Нотінгемська база народної музики» [11].

Порівняння значень функції витрат при 200 епохах наведених в табл. 1.

ТАБЛИЦЯ 1 Порівняння характеристик топологій RHM-ОМБ, ДКЧП-ОМБ, RHM-ГМП, ДКЧП-ГМП

Якісні характеристики	Структура топології			
	RHM-ОМБ	ДКЧП-ОМБ	RHM-ГМП	ДКЧП-ГМП
NLL	-4,73434	-4,68098	-3,80324	-3,71452
CEC	-0,2409	-0,2367	-0,2241	-0,2056
ACC, %	73,2	76,1	77,4	79,8
Час тренування, годин	5-7	8-9	12-13	16-18

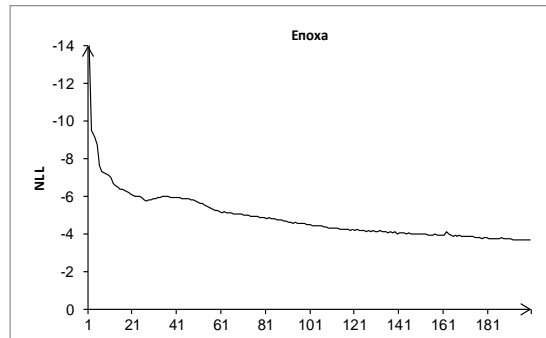


Рис. 8 Графік падіння цільової функції ДКЧП-ГМП

Графіки падіння цільової функції при різних топологіях наведені на рис. 5 – рис. 8.

A. Переваги та недоліки

Провівши низку дослідів та експериментів, можна виділити основні недоліки та переваги топологій RHM-ОМБ, ДКЧП-ОМБ, RHM-ГМП, ДКЧП-ГМП.

Переваги топології RHM-ОМБ:

- час тренування та генерації, у порівнянні з іншими топологіями, в даному підході до генерації музики тривав найменшу кількість годин, що підтверджено експериментально;
- економічні витрати – виконання цього алгоритму не потребує великих обчислювальних потужностей.

Недоліки топології RHM-ОМБ:

- неточний ймовірнісний розподіл у порівнянні з глибинними моделями;
- недоліки RHM – проблема зникаючого градієнту.

Переваги топології ДКЧП-ОМБ:

- час тренування дещо більший, ніж у RHM-ОМБ, але все ще менший у порівнянні з глибинними топологіями, що підтверджено експериментально;
- переваги ДКЧП – вирішення проблеми довгострокових залежностей.

Недоліки топології ДКЧП-ОМБ:



- реалізація – складна будова вузла рекурентної частини;
- економічні витрати – необхідність використання більшої кількості обчислювальних ресурсів, на яких виконується тренування, що пояснюється в застосуванні формул та виразів для обчислення значення виходу поточного вузла ДКЧП, що призводить до підвищення кількості необхідних обрахунків.

Переваги топології РНМ-ГМП:

- можливість відтворення глибоких залежностей та зв'язків між нотами;
- точність вища, а значення функції правдоподібності менше, у порівнянні з моделями на основі обмежених машин Больцмана.

Недоліки топології РНМ-ГМП:

- недоліки РНМ – проблема зникаючого градієнту.
- економічні витрати – необхідність в підвищенні продуктивності обчислювальної машини, на якій виконується тренування, що пояснюється збільшенням кількості необхідних обчислень, а саме ймовірного розподілу по всіх прихованих шарах.

Переваги ДКЧП-ГМП:

- точність – в порівнянні з вищеописаними моделями ця топологія показує найбільшу точність, що досягається поєднанням особливостей її частин, а саме можливістю обробки довгострокових залежностей та глибинного представлення даних;
- переваги ДКЧП – вирішення проблеми довгострокових залежностей.
- значення функції втрат є найменшим з усіх топологій.

Недоліки ДКЧП-ГМП:

- час тренування – серед описаних вище топологій даний підхід вимагає найбільше часу, що пов'язано з великою кількістю обрахунків;
- економічні витрати – необхідність в підвищенні продуктивності обчислювальної машини, на якій виконується тренування, що також пояснюється збільшенням кількості необхідних обчислень.

В. Вибір топології

Результати порівняння показують, що серед топологій РНМ-ОМБ, ДКЧП-ОМБ, РНМ-ГМП та ДКЧП-ГМП найкращі показники якості має ДКЧП-ГМП, проте час тренування на невеликому дата-сеті зайняв 16 годин. Водночас найшвидшою виявилася топологія РНМ-ОМБ, проте її показники якості найгірші серед порівнюваних. Можна звернути увагу на те, що найбільша різниця серед значень показників якості

спостерігається при переході від ОМБ до ГМП, що і відбивається на часі тренування. Оскільки обрана топологія має бути основою генератора музики, час тренування на нових вибірках та дата-сетах має вирішальне значення, оскільки показники при таких значеннях не впливають на якість згенерованої музики і різниця є суб'єктивною та відрізняється тільки кількістю можливих зв'язків нот, що можуть бути згенеровані.

Таким чином, за основу генератора музики була обрана ДКЧП-ОМБ, що в порівнянні з РНМ-ОМБ вирішує проблему довгострокових залежностей. Водночас, використовуючи більш продуктивне обладнання, ніж те, на якому проводилося дослідження, можна застосувати ДКЧП-ГМП, взявши до уваги, що її показники якості, а саме NLL, на 27% менші за NLL РНМ-ОМБ.

ВИСНОВКИ

В результаті проведення цього дослідження було порівняно різні топології нейронних мереж, що можуть бути застосовані до вирішення завдання генерації музики. Результати показують, що найкращі показники дала топологія ДКЧП-ГМП, проте час тренування на невеликому дата-сеті склав 16 годин. Тому, беручи до уваги, що для реалізації генератора музики час тренування має вирішальне значення, за основу була обрана ДКЧП-ОМБ, яка в порівнянні з РНМ-ОМБ вирішує проблему довгострокових залежностей та має кращі показники якості. Генератор на основі ДКЧП-ОМБ створює досить «прийнятну» музику, яка є поліфонічною та має невелику кількість випадкових нот, тобто були дотримані всі вимоги. Однак «прийнятність» створеної музики є суб'єктивною оцінкою, а якість згенерованої музики не оцінювалась суб'єктивними методами, що робить це основним недоліком дослідження, яке планується усунути в майбутньому шляхом голосування та зіставлення показників якості, обрахованих математично, з результатами голосування, що дає змогу спрогнозувати суб'єктивну оцінку з певною ймовірністю.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] E. Douglas, "Make Music and Art Using Machine Learning," *Magenta*, 2016. [Online]. Available: <https://magenta.tensorflow.org/>.
- [2] A. van den Oord, S. Dieleman, and H. Zen, "WaveNet: A Generative Model for Raw Audio," *DeepMind*, 2016. [Online]. Available: <https://deepmind.com/blog/wavenet-generative-model-raw-audio/>.
- [3] D. Johnson, "Composing Music With Recurrent Neural Networks · hexahedria," *hexahedria*, 2015. [Online]. Available: <http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/>.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, "Sequence Modeling: Recurrent and Recursive Nets," in *Deep Learning*, MIT Press, 2016, pp. 367–415, URL: <http://www.deeplearningbook.org/contents/rnn.html>.
- [5] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies," 1991, URL: <http://www.bioinf.jku.at/publications/older/ch7.pdf>.



- [6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [7] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," 2015, arXiv: [1506.00019](https://arxiv.org/abs/1506.00019).
- [8] G. (LISA lab. . Hinton, "Restricted Boltzmann Machines (RBM)," *DeepLearning 0.1 documentation*, 2010. [Online]. Available: <http://deeplearning.net/tutorial/rbm.html>.
- [9] G. Hinton, "Deep Belief Nets," Toronto, 2007, URL: <https://www.cs.toronto.edu/~hinton/nipstutorial/nipstut3.pdf>.
- [10] N. Boulanger-Lewandowski, "Modeling High-Dimensional Audio Sequences with Recurrent Neural Networks," Universit_e de Montr_eal, 2014. URL: http://www-etud.iro.umontreal.ca/~boulanni/NicolasBoulangerLewandowski_thesis.pdf
- [11] J. Allwright, "ABC version of the Nottingham Music Database," 2003. [Online]. Available: <http://abc.sourceforge.net/NMD/>.

Надійшла до редакції 24 червня 2017 р.

УДК 004.891.2

Применение нейронных сетей для решения задачи генерации музыки

Дорогой Я. Ю., к.т.н., доц., ORCID [0000-0003-3848-9852](https://orcid.org/0000-0003-3848-9852)

e-mail argusyk@gmail.com

Цуркан В. В., к.т.н., ORCID [0000-0003-1352-042X](https://orcid.org/0000-0003-1352-042X)

e-mail v.v.tsurkan@gmail.com

Бельдяга К. Д. ORCID [0000-0002-6456-6981](https://orcid.org/0000-0002-6456-6981)

e-mail kirill.beldyaga@gmail.com

Национальный технический университет Украины

"Киевский политехнический институт имени Игоря Сикорского" kpi.ua

Киев, Украина

Реферат—В статье рассматривается возможность использования нейронных сетей в качестве основы генератора музыки. Под генерацией музыки понимают процесс создания последовательностей музыкальных нот, относя данную задачу к задачам моделирования последовательностей, как и моделирование речи. Актуальность развития данной темы связана с необходимостью написания новой музыки и нахождения новых мотивов, что вызвано производством большого количества фильмов и потребностями людей. Основным требованием к системе была генерация полифонической музыки, которая является последовательностью кортежей нот, то есть многомерным объектом, что требует использования порождающих моделей, которые способны вычислять вероятностное распределение по отношению к рекуррентным моделям, которые, в свою очередь, позволяют воссоздать долгосрочные зависимости в последовательности многомерных объектов. Для этого в статье приведены необходимые теоретические сведения. Рассмотрены существующие решения и определены их преимущества и недостатки. Проанализированы и представлены варианты возможных топологий, алгоритмы их тренировки и генерации. Описаны результаты экспериментальных исследований, на основании которых выполнено сравнение и определены преимущества и недостатки каждой из них.

Библ. 11, рис. 8, табл. 1.

Ключевые слова — нейронные сети; музыка; моделирование последовательностей; генерация музыки; машинное обучение.



Application of Neural Networks to Solve the Task of Generating Music

Ya. Yu Dorohyi, PhD, Assoc.Prof., ORCID [0000-0003-3848-9852](https://orcid.org/0000-0003-3848-9852)

e-mail argusyk@gmail.com

V. V. Tsurkan, PhD, ORCID [0000-0003-1352-042X](https://orcid.org/0000-0003-1352-042X)

e-mail v.v.tsurkan@gmail.com

K. D. Beldiaha, ORCID [0000-0002-6456-6981](https://orcid.org/0000-0002-6456-6981)

e-mail kirill.beldyaga@gmail.com

National technical university of Ukraine "Igor Sikorsky Kyiv polytechnic institute" kpi.ua
Kyiv, Ukraine

Abstract— The article considers the possibility of using neural networks as the basis of the music generator. Under the music generation is the process of creating sequences of music notes, which relates this task to the tasks of simulation sequences, as well as modeling the language. The urgency of the development of this topic is connected with the necessity of writing new music and finding new motives, caused by the production of many films and people's needs. The main requirement for the system was the generation of polyphonic music, a sequence of musical tuples, that is, a multi-dimensional object requiring the use of generative models, such as restricted Boltzman machine, that are able to calculate the probabilistic distribution, by estimating the energy function of configuration state, in relation to recurrent patterns, which in turn enable their production of long-term dependencies in a sequence of multi-dimensional objects. And for the representing of deep relations and dependencies between music notes there were deep belief network used that is able to estimate deep connections and dependencies by hierarchical representing of data. This approach allows to create four topologies which are compound of recurrent part and generative. For this purpose in the article necessary the or ethical information is given. The existing solutions are considered and the iradvantages and disadvantages are determined. The variants of possible topologies, algorithms of the ir training and generation are analyzed and presented. The results of experimental researches on the basis of which the comparisonis made and the advantages and disadvantages of each of the mare determined. For the comparison there were used the negative log-likelihood function, cross-entropy cost, accuracy and training time. The results show that the best characteristics has topology that consists of long short-term memory and deep belief network, because of it difficult structure and properties. But at the same time, training time of this topology was extremely high on small dataset, that makes not efficient to use it as a basis of music generator. Topology which consists of recurrent neural network and restricted Boltzman machine had the worst indicators, but the small training time. Thereby topology which consists of long short-term memory and restricted Boltzman machine had become the best candidate. Good characteristics of this structure are confirmed by the main property of long short-term memory that is solution of vanishing gradient problem and by the possibility of representing complicated distribution in restricted Boltzman machine. Training were performed on "Nottingham folk music database" dataset through 200 epochs. All training and generating algorithms were implemented in Python with Theano. Research described in this article allows to simplify the process of development of polyphonic music generator.

Ref. 11, fig. 8, tabl. 1.

Keywords — *neuralnetworks; music; modelingsequences; musicgeneration; machinelearning.*

