

О.С. Комарський, А.Ю. Дорошенко

## МОДЕЛЬ РЕКУРЕНТНОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ГЕНЕРАЦІЇ МУЗИКИ

У роботі розглядається можливість генерації музичних композицій, використовуючи рекурентні нейронні мережі. Запропоновано та розглянуто два методи генерації музичних творів – на рівні нот та на рівні акордів. Проведено дослідження обох методів та визначено їх переваги та недоліки. Далі для розробки обрано метод генерації на рівні нот, для якого детально описано процес пошуку та обробки даних для навчання генерації музики за допомогою рекурентної нейронної мережі. Це дозволяє автоматизувати генерацію музичних творів без втручання людини. Для побудованої моделі виконано програмну реалізацію запропонованого рішення, проведені експерименти та їх верифікація за участю фокус-групи людей щодо визначення авторства створеної музики – людина чи комп'ютер.

Ключові слова: генерація музики, машинне навчання, рекурентні нейронні мережі, LSTM, RNN

### Вступ

Машинне навчання покращує багато аспектів нашого повсякденного життя, як очевидних, так і непомітних. Штучний інтелект має велику вагу у таких речах, як системи рекомендацій в різних медіа-сервісах, розумні будинки, комп'ютерний зір, системи виявлення підозрілих дій тощо. І, незважаючи на те, що досі триває дискусія навколо нейронних мереж як чорної скриньки – існує величезний потенціал для їх використання в різних сферах людської діяльності, включно із медициною, віртуальними помічниками, електронною комерцією та сферою фінансів.

Водночас не менш важливий та цікавий напрям, в якому машинне навчання може відіграти свою роль – це творчість. Комп'ютерна творчість є дійсно захоплюючою сферою, в якій штучний інтелект перетинається з мистецтвом, а метою є розроблення систем, які здатні моделювати, вдосконалювати або розуміти творчість людини.

Генерація музики є найбільш абстрактним напрямком творчої діяльності, оскільки на відміну від написання книжок, картин або віршів, музика не прив'язана до контексту оточуючого нас світу. Водночас запис музики може бути представлений у вигляді дуже простого символічного формату даних без втрати змісту.

Наше життя тісно пов'язане з музикою. Ми отримуємо задоволення від прослуховування музичних композицій, а

також використовуємо її у відеороликах, кіно, презентаціях та рекламі. Тому здатність генерувати музику високої якості є значним кроком у житті людини. Проте досі немає відповіді на питання, чому одну послідовність нот люди сприймають як музику, а іншу ні. Це свідчить про те, що в музиці існують деякі інтуїтивно зрозумілі всім закономірності, які досі не вдалося математично строго сформулювати.

Існують два основні напрями розробки в сфері генерації музики. Перший вирішує задачу генерації музичних композицій без участі людини. Другий – спрямований на створення програмного забезпечення, яке допомагає людині створювати власні композиції. Розглядаючи детальніше тематику генеративного мистецтва, можна дійти висновку, що програмне забезпечення для автоматичної генерації музичних композицій є дуже перспективною розробкою.

Основною метою даної роботи є створення системи для автоматичної генерації музики, використовуючи рекурентні нейронні мережі.

### 1. Генеративне мистецтво

Автоматична генерація музики – це процес створення музичного твору з мінімальним втручанням людини.

Алгоритмічний підхід до створення музики існує вже кілька століть, починаючи з Гвідо д'Ареццо, який 1024 року

винайшов перший алгоритм для складання музики [1]. Не зважаючи на те, що в докомп'ютерну епоху існувало кілька підходів до алгоритмічного створення музики, найяскравіші результати були отримані лише з появою комп'ютерів. Саме через величезні можливості, які пропонує обчислювальна техніка, алгоритмічні музичні композиції почали розквітати від початку 1950-х років та до сьогодні [2].

### 2. Модель нейронної мережі

Для побудови якісної нейронної мережі необхідно обрати відповідну базову архітектуру нейронної мережі [3]. Було вирішено обрати архітектуру AWD-LSTM, оскільки ця модель часто використовується для побудови нейронних мереж, які працюють у напрямку мовних досліджень [4].

Хоча багато моделей для генерації музики використовують невеликий діапазон нот, у даній реалізації модель використовує діапазон із 62 нот (цей діапазон охоплює більшість класичної музики) та дозволяє відтворювати будь-яку кількість нот та інструментів одночасно. Для навчання було використано великий датасет MIDI файлів із класичною музикою. Але, оскільки в моделі немає ніяких специфічних налаштувань для класичної музики, то можна використовувати датасет будь-якого музичного жанру.

### 3. Методи нот та акордів

Одним із можливих рішень для генерації музики – щоразу запитувати модель, чи грати їй зараз цю ноту для кожної з 88 клавіш фортепіано на кожному музичному кроці. Однак, оскільки кожна окрема нота здебільшого мовчить, то для нейронної мережі важко було б навчитись передбачати таку ситуацію для нот. Тому варто розглянути два можливі способи для рішення цієї проблеми.

Мовні нейронні мережі часто працюють на рівні символів, або на рівні слів. Так само для генерації музики, тут досліджується два аналогічних рівня – нот та акордів.

Акордом вважається кожна комбінація нот, які будь-коли зустрічалися в музичній композиції. Тобто сприймати акорд

можна так само, як слова. Це означає, що можливо запитати нейронну мережу, яким буде наступний акорд (слово), надавши їй попередньої послідовності акордів (речення). Потрібно розуміти, що мається на увазі акорд, як будь-яка комбінація нот, що відтворюються одночасно. Не обов'язково це мають бути традиційні музичні акорди.

Для методу, який використовує ноти, вважаємо, що початок та кінець кожної ноти є різними словами. Наприклад «p28» та «endp28» у текстовому форматі MIDI є різними словами, та будуть використовуватись для початку та зупинки ноти №28.

В обох методах було використано дещо зменшений діапазон нот – 62 ноти замість повних 88 на фортепіано. Ноти які виходили з цього діапазону, було переміщено вниз або вгору на октаву.

В результаті експериментів були визначені характеристики кожного з двох методів.

Характеристика методу акордів:

1. Нейронна мережа добре генерує довгі музичні патерни, але не може вийти за межі навчальних даних та згенерувати щось нове.

2. Зазвичай для прийняття рішення щодо наступного акорду, найефективніше обирати найбільш імовірний прогноз на кожному часовому відрізку.

3. В результатах зустрічається дивний ефект, коли час від часу музична композиція перестрибує на іншу неочікувану ноту, але потім одразу продовжує грати мелодію в звичайному ритмі.

Характеристика методу нот:

1. Нейронна мережа створює гармонійні ритмічні малюнки, які можуть легко переходити один в другий.

2. Легко обробляє параметри тривалості натискання ноти, через що мелодія здається більш інтуїтивно зрозумілою для людини.

3. Зазвичай для прийняття рішення щодо наступної ноти найефективніше чергувати між собою три найбільш імовірні прогнози на кожному часовому відрізку.

Отже, в результаті цих порівнянь було прийнято рішення про використання методу нот як основного для генерації музики в системі.

#### 4. Навчання нейронної мережі

Навчання моделі нейронної мережі потребує багато обчислювальних ресурсів та велику кількість якісних навчальних даних. Але, чим більше даних буде використовуватися, тим більше часу та обчислювальної потужності необхідно моделі для завершення навчання [5].

Для прискорення навчання були використані хмарні обчислення за допомогою ресурсів графічного процесора, оскільки такий підхід забезпечує більшу швидкість опрацювання даних, ніж центральний процесор. Це видно з результатів порівняння, зображених на рис. 1. Для порівняння були використані центральний процесор i7 8550 U на 4 ядра, та графічний процесор Geforce MX150 з 2 гігабайтами оперативної пам'яті. Як видно з діаграми, із зростанням кількості даних зростала швидкість їх обробки, особливо використовуючи графічний процесор, який за максимальної кількості даних був швидшим, аніж центральний процесор у 5,5 разів.

Вибір даних для тренування є важливою складовою будь-якої нейронної мережі, оскільки вона отримує всі свої «знання» про музику лише з навчальних даних. Тому, чим більше буде даних для навчання, та чим якісніше вони будуть, тим кращими будуть результати, згенеровані нейронною мережею. Також необхідно мати дані для тестування нейронної мережі, які повинні відрізнятися від даних для тренування, щоб визначати, чи змогла модель абстрагуватися

від даних тренування та скласти нові мелодії, а не просто відтворювати дані.

До тренувальних даних є дві вимоги. Перша – музичні композиції мають бути в одному жанрі та написані людьми, а не іншою нейронною мережею. Друга вимога – мелодії мають приємно звучати. Оскільки зазвичай різні жанри музики для людей звучать по-різному, адже залежать від їхнього смаку, то було вирішено обрати для навчання класичну музику як найбільш нейтральний жанр. До того ж класична музика – це найдоступніший музичний жанр у форматі MIDI.

Для збору навчального датасету було написано декілька скриптів, за допомогою яких були зібрані файли з найбільшого сайту агрегатора класичної музики – «Classical Piano Midi Page», де представлені композиції таких композиторів як Моцарт, Бах, Бетховен та багато інших.

Для навчання моделі необхідно було перетворити зібрані MIDI-файли у формат, який зможе зрозуміти нейронна мережа. Файли MIDI надають інформацію про час початку гри та зупинки на кожную ноту, а також інформацію про музичний інструмент, значення гучності та темпу гри [6]. Але, оскільки цей формат має специфічні позначення, які не дуже зручно прогнозувати, було написано функцію, котра може перекодувати їх до зручнішого текстового формату, який може зрозуміти модель нейронної мережі. Також на виході нейронної мережі

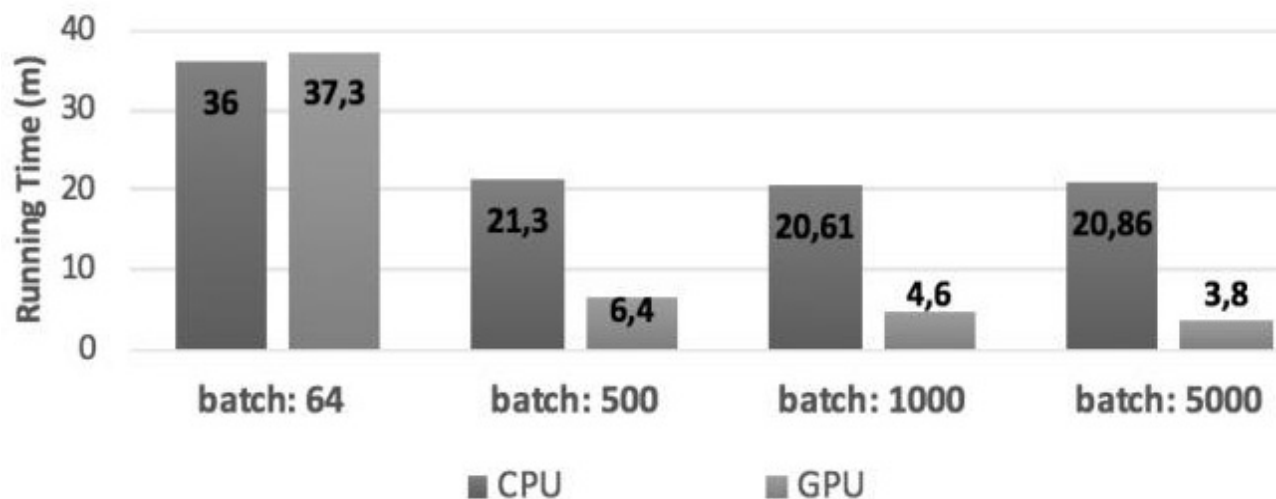


Рисунок 1. Порівняння швидкості навчання нейронної мережі з використанням CPU та GPU

музичні композиції подаються спочатку у вигляді власного текстового формату, а далі кодуються в форматі MIDI та MP3 (рис. 2).

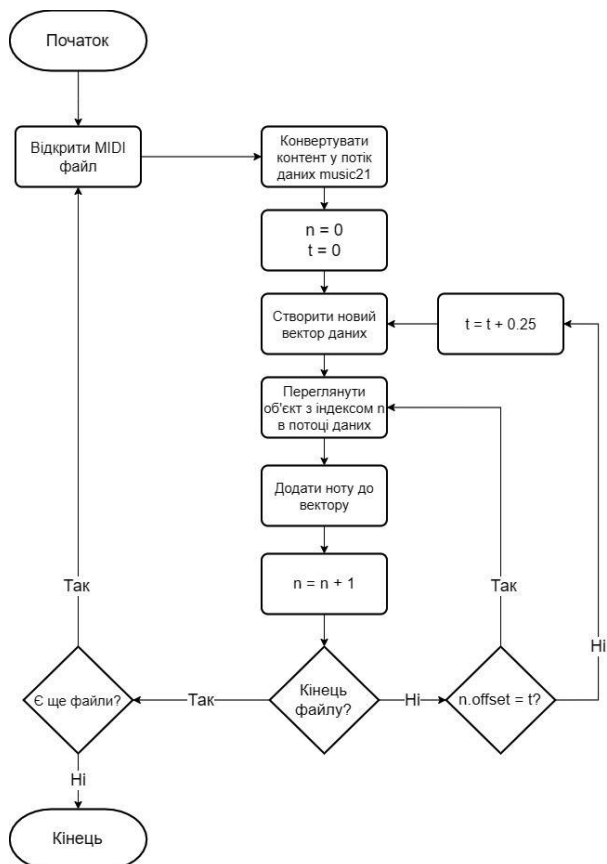


Рисунок 2. Блок-схема алгоритму перетворення формату MIDI у текст.

### 5.Результати

Під час тестування роботи нейронної мережі було досліджено вплив зміни деяких параметрів нейронної мережі на функцію втрати [4].

Експериментальним шляхом було визначено, що оптимальними характеристиками рекурентної нейронної мережі для генерації музики є мережа, яка складається з 4 LSTM шарів [4], кожен розмірністю по 600 нейронів зі значенням показника виключення (dropout) = 1 та кількістю епох 150.

Фінальний графік функції втрати нейронної мережі з обраними характеристиками зображено на рис. 3.

### 6.Оцінка якості нейронної мережі

Через те, що основною задачею є генерація музичних творів, то її якість не можна оцінити за об'єктивними характеристиками, оскільки оцінка краси, або естетичної цінності у творах мистецтва зводиться до індивідуальної суб'єктивної думки людей.

Для цього було розроблено спеціальне опитування, метою якого є оцінка наступних цілей, поставлених перед результатами генерації:

- музичні композиції повинні приємно та гармонійно звучати для людського вуха;

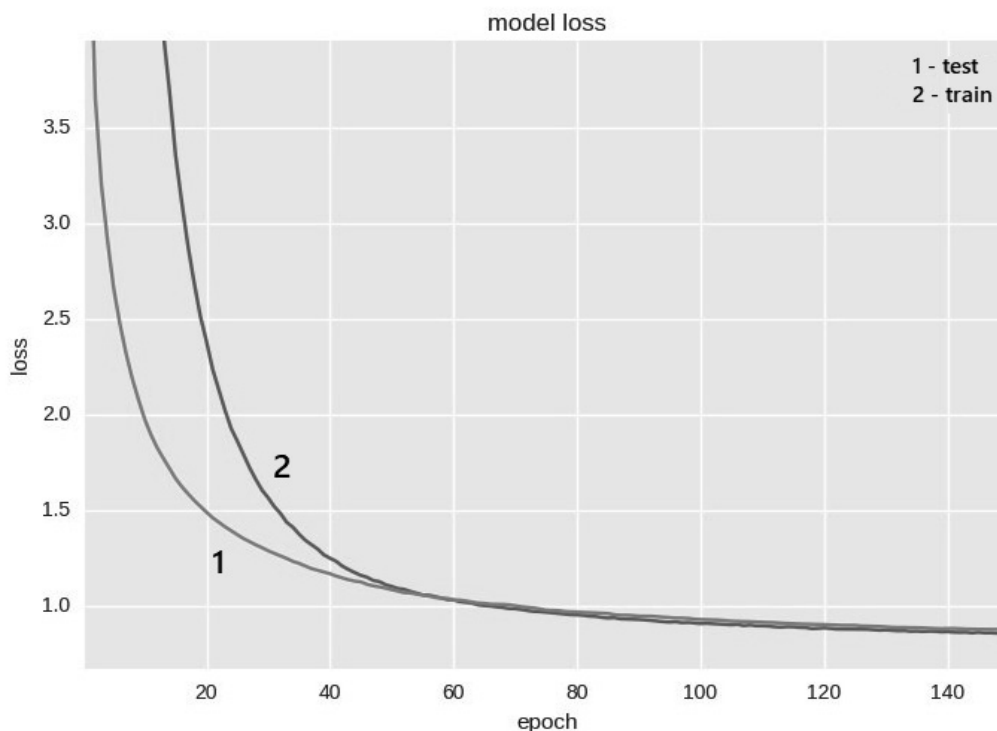


Рисунок 3. Графік функції втрати для фінальної моделі нейронної мережі.

– при прослуховуванні не має бути зрозумілим, чи мелодію було згенеровано нейронною мережею, чи людиною.

Для суб'єктивної оцінки якості музики згенерованої нейронною мережею, було підготовлено опитування, створене на платформі Google Forms. Тест було розроблено так, щоб можна було чітко зробити висновки та відповіді на два основних запитання про якість музичних композицій.

На початку проходження тесту для більш чіткої обробки результатів визначається група, до якої належить людина (фокус-група), яка буде проходити тест. Для відповіді було запропоновано декілька варіантів основних груп, що мають зміст у рамках тесту.

Наступним етапом є безпосереднє опитування щодо якості музичних композицій. Для цього спочатку надається можливість прослухати композицію, після чого пропонується відповіді на два запитання. Перше запитання стосується якості звуку, тобто суб'єктивної оцінки приємності звучання за шкалою від одного до десяти, де 1 – погано, 10 – дуже добре. Друге запитання стосується визначення автора твору – комп'ютер або людина. Усього 6 подібних етапів для 6 музичних творів.

Кожна музична композиція, яка наведена в опитуванні, – це, або згенерована музика нейронною мережею, що створена в рамках цієї роботи, або одна з компози-

цій написаних людиною, яка була взята з навчального набору даних.

Також на початку опитування учасникам повідомляється, що принаймні одна з композицій була створена комп'ютером, а також принаймні одна є справжнім музичним твором, написаним одним із відомих композиторів.

Отже, опитування є також своєрідним «музичним тестом Тюринга», оскільки фокус-група одночасно оцінює якість творів, а також визначає, створений він людиною чи комп'ютером.

Опитування проводилось протягом тижня, за цей час у ньому взяли участь 55 людей. Опитування відбувалося в місцях, де потенційно можуть знаходитися люди з цільових груп, тобто в різних музичних спільнотах.

Переважно це були люди, які люблять слухати музику (43,6%), але не мають досвіду гри на музичних інструментах. На другому місці за кількістю людей у фокус-групі були ті, хто мають певний досвід гри на інструментах (32,7%). Решта – це учасники з музичною освітою (16,4%), або професійні музиканти, які постійно мають справу з музикою у роботі (7,3%). Детальний розподіл учасників опитування за музичним досвідом можна побачити на рис. 4.

Також були підраховані середні значення оцінки якості музичних композицій



Рисунок 4. Розподіл учасників опитування за їх музичним досвідом.

за результатами опитування, які наведені на рис. 5.

Отже, як бачимо з результатів опитування, загальна середня оцінка за трьома композиціями майже однакова, але дещо більша у творів написаних людиною. Проте, оскільки середні результати по кожній композиції майже на одному рівні, це означає, що мелодії, згенеровані нейронною мережею є якісними, оскільки подобаються фокус-групі так само, як і композиції, написані людиною. Також жодна з композицій не отримала оцінки нижче 3.

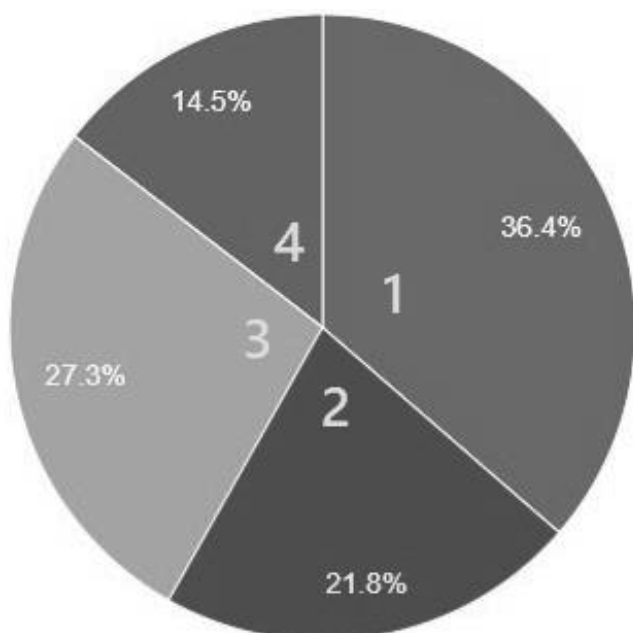
Наступним етапом обробки результатів опитування є підрахунок відповідей

щодо визначення авторства композиції – людина чи комп'ютер. На рис. 6 бачимо зведену діаграму, яка показує у відсотках випадки, коли учасники опитування вірно або невірно визначали автора.

Отже, учасникам опитування було важко правильно визначити, хто є насправді автором прослуханих музичних творів – людина чи нейронна мережа. Згідно з діаграмою, на якій зведені результати опитування на це запитання, фокус-групі вдалося правильно відповісти лише в 58,2% відповідей, що трохи краще, ніж звичайний шанс випадковості на запитання з двома варіантами відповіді (50%). Також ціка-

Група	Композиції згенеровані нейронною мережею			Композиції створені людиною		
	Перша	Друга	Третя	Перша	Друга	Третя
Середня оцінка якості	5,54	6,62	6,12	6,32	7,12	5,5
Загальна оцінка якості за групою	6,09			6,31		

Рисунок 5. Середні показники якості музичних композицій за результатами опитування.



- 1 - Обрали комп'ютер.  
Автор комп'ютер
- 2 - Обрали людина.  
Автор людина
- 3 - Обрали людина.  
Автор комп'ютер
- 4 - Обрали комп'ютер.  
Автор людина

Рисунок 6. Результати опитування на питання про визначення автора музичного твору.

во, що 27,3 % опитуваних хибно вважали, музичний твір, створеним людиною, тоді як насправді він був згенерований нашою системою. Це свідчить про те, що учасникам було важко відрізнити згенеровані твори від написаних композиторами.

### Висновки

У роботі була розглянута й реалізована можливість використання рекурентних нейронних мереж для задачі генеративного мистецтва. Були запропоновані та розглянуті два підходи, які використовують метод нот та метод акордів. У результаті досліджень основним був обраний метод нот, оскільки результатом його роботи були цікавіші та гармонійніші музичні композиції, на відміну від результатів, які вдалось отримати методом акордів. Також було здійснено порівняння швидкості навчання нейронної мережі з використанням центрального та графічного процесорів, за результатами якого графічний процесор обробляв дані швидше за центральний у 5,5 разів.

Для оцінки якості генерації музичних творів було проведено опитування фокус-групи, яке показало, що музика згенерована нейронною мережею, отримала майже такі ж оцінки, як і музика, написана людиною, що є відмінним результатом. Також було встановлено, що учасникам опитування було важко правильно визначити автора музичного твору, адже вони вірно визначали авторів лише в 58 % випадків.

### References

1. *Nierhaus G.* (2009) *Algorithmic Composition - Paradigms of Music Generation.* Springer, Vienna. pp. 7-66.
2. *Wasserman P.* (1992) *Neural Computing : Theory and Practice.* pp. 180-185.
3. *Callan R.* (2001) *The Essence of Neural Networks.* pp. 50-65.
4. *Nielsen A.* (2015) *Neural Networks and Deep Learning.* p. 111.

5. *Kandel E., Schwartz J., Jessell T.* (2000) *Principles of Neural Science 4th Edition.* pp. 283-302.
6. *Introduction to MIDI and Computer Music.* URL:<https://cecm.indiana.edu/361/midi.html>

Отримано: 17.02.2022

### Про авторів:

Комарський Олександр Сергійович, магістрант Національного технічного університету України «КПІ імені Ігоря Сікорського».  
<https://orcid.org/0000-0002-1005-6863>

Дорошенко Анатолій Юхимович, доктор фізико-математичних наук, професор кафедри інформаційних систем та технологій Національного технічного університету України «КПІ імені Ігоря Сікорського» та завідувач відділу теорії комп'ютерних обчислень Інституту програмних систем НАН України. Кількість наукових публікацій в українських виданнях – понад 190. Кількість наукових публікацій в зарубіжних виданнях – понад 80. Індекс Хірша – 6.  
<http://orcid.org/0000-0002-8435-1451>

### Місце роботи авторів:

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», проспект Перемоги 37 та Інститут програмних систем НАН України, 03187, м. Київ-187, проспект Академіка Глушкова, 40. Тел.: (044) 526 3559  
E-mail: [komarskiy33@gmail.com](mailto:komarskiy33@gmail.com),  
[doroshenkoanatoliy2@gmail.com](mailto:doroshenkoanatoliy2@gmail.com)