

СИНЕГЛАЗОВ Віктор Михайлович,

доктор технічних наук, професор,

Державний університет «Київський авіаційний інститут», Київ, Україна,

<https://orcid.org/0000-0002-3297-9060>

ЧЕЛКОВАН Віктор Іванович,

Центральний науково-дослідний інститут Збройних Сил України, Київ, Україна,

<https://orcid.org/0009-0001-4522-6260>

МОДИФІКОВАНИЙ ІЄРАРХІЧНИЙ ТРАНСФОРМЕР ДЛЯ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ВІЙСЬКОВОГО ПРИЗНАЧЕННЯ

Мета статті. Побудова модифікованого ієрархічного трансформера (архітектури глибокого навчання нейронної мережі, що побудована на механізмі самоуваги) завдяки додаванню когнітивних карт контексту для покращення процесу оброблення різномірної інформації в інтелектуальних системах підтримки прийняття рішень військового призначення.

Методи дослідження. Під час дослідження було застосовано: метод аналізу механізмів роботи уваги в трансформерах, метод математичного моделювання взаємодії об'єктів у бойовому просторі, а також метод формалізації когнітивних карт на основі концепцій з нейронауки. Ці методи дали змогу теоретично обґрунтувати побудову ієрархічного трансформера з інтегрованими когнітивними картами контексту.

Отримані результати дослідження. Сформульовано теоретичні підходи до модифікації ієрархічного трансформера з інтегрованими когнітивними картами контексту. Запропоновано формальну математичну модель когнітивної карти контексту як структури, що описує множину станів тактичної (оперативної) обстановки, причинно-наслідкові зв'язки між ними, їх векторне подання та динаміку переходів. Розроблено модифікований механізм самоуваги (self-attention) з інтеграцією матриці когнітивних впливів. Обґрунтовано ієрархічну організацію когнітивних карт контексту з механізмом узгодження розмірностей просторів між рівнями. Запропоновано механізм темпоральної синхронізації для обробки асинхронних потоків даних.

Елементи наукової новизни. Вперше запропоновано модифікацію механізму самоуваги трансформерів через інтеграцію формалізованих когнітивних карт контексту. Такий підхід забезпечує інтерпретованість прийнятих рішень через явне подання причинно-наслідкових зв'язків між елементами бойової обстановки, а також дає змогу однотипно обробити різномірну інформацію: геопросторові координати об'єктів, темпоральні залежності з ефектами згасання впливу різних факторів та дискретних подій на об'єкти у часі. Для ієрархічної організації когнітивних карт запропоновано застосувати механізм узгодження розмірностей просторів між різними рівнями для забезпечення просторово-контекстуальної агрегації інформації.

Теоретичне та практичне значення викладеного у статті зводиться до створення математичного та алгоритмічного фундаменту для розробки інтелектуальних систем підтримки прийняття рішень військового призначення. Одним із основних призначень таких систем є оброблення гетерогенних інформаційних потоків різних джерел у реальному часі. Запропонований трансформер забезпечує високий рівень інтерпретованості рішень через механізм декомпозиції тактичної (оперативної) обстановки. Ця модифікована архітектура з інтегрованими когнітивними картами забезпечує розуміння командуванням причинно-наслідкових зв'язків між елементами обстановки та запропонованими системою рішеннями. Запропонований модифікований трансформер з інтегрованими когнітивними картами забезпечує ефективну обробку різномірної інформації через модифікацію механізму самоуваги, ієрархічну організацію когнітивних карт та темпоральну синхронізацію асинхронних інформаційних потоків. Інтеграція експертних знань через матрицю когнітивних впливів забезпечує інтерпретованість рішень і можливість багатокритеріальної декомпозиції. Перспективами подальших досліджень є: експериментальна валідація запропонованого трансформера на реальних або синтетичних датасетах бойових тактичних (оперативних) епізодів, а також розроблення методів автоматизованого вилучення когнітивних карт контексту з історичних даних.

Ключові слова: трансформери, когнітивні карти, системи підтримки прийняття рішень, мультимодальна обробка даних, механізм самоуваги, штучний інтелект, ієрархічна інтеграція даних, темпоральна синхронізація.

Вступ

Постановка проблеми. Під час ведення бойових дій в сучасних умовах особовий склад сил безпеки і оборони України, від солдата до генерала, має справу з великим обсягом даних різного характеру. Розвідувальні дані, метеорологічні умови, дані про місцевість, стан і положення своїх військ та противника мають різну динаміку оновлення, формати подання та важливість для прийняття рішень. Чи здатні традиційні підходи забезпечити коректну семантичну інтеграцію даних? Практика свідчить, що ні. Це призводить до неповної або суперечливої картини оперативної обстановки [1].

Наразі системи підтримки прийняття рішень переважно реалізують алгоритми та логічні правила, що виконують фіксовану послідовність дій для окремих етапів процесу прийняття рішень та мають суттєві обмеження. Наприклад, приведення різнорідної інформації до уніфікованого табличного формату призводить до втрати важливих контекстних зв'язків. Це і просторові відношення між об'єктами, часові залежності, тактичний вплив одних об'єктів на інші. Традиційні підходи не здатні ефективно моделювати складні нелінійні залежності між різними типами даних та їх впливом на оперативну обстановку [2]. На відміну від них, системи на основі глибокого навчання здатні адаптуватися до нових ситуацій та виявляти приховані закономірності в даних. Для обробки послідовностей даних використовуються рекурентні нейронні мережі (RNN), згорткові мережі (CNN) й трансформери.

Так, *рекурентні мережі* обробляють дані послідовно. Це обмежує їх можливості паралельних обчислень, а також ускладнює обробку довгих залежностей [3]. *Згорткові мережі* добре обробляють локальні патерни. Проте обробка довгих послідовностей з глобальними взаємозв'язками не є ефективною [4]. *Трансформер* – це архітектура глибокого навчання нейронної мережі, побудована на механізмі самоуваги. Трансформери як архітектура здійснили прорив у такому технологічному напрямі штучного інтелекту як обробка природної мови [5]. Механізм самоуваги дає змогу обчислювати релевантність вхідної інформації для конкретного завдання. Цей механізм визначає ваги взаємозв'язків між усіма парами елементів вхідної послідовності. Для кожного елемента формуються три вектори: запит (query), ключ (key) та значення (value). Вага зв'язку між елементами визначається скалярним добутком їх запитів і ключів, після чого використовується нормалізація. Така властивість дає змогу враховувати залежності між усіма елементами незалежно від їх відстані у послідовності.

Проте пряме застосування стандартних трансформерів у військовій сфері стикається з критичними викликами. Наприклад, квадратична обчислювальна складність, яка унеможливує оброблення великих обсягів даних у реальному часі. Також, через відсутність механізмів інтеграції експертних знань – неможливість забезпечити необхідний рівень інтерпретованості рішень для їх валідації військовими експертами [6]. Додатковою

проблемою є відсутність механізмів урахування причинно-наслідкових зв'язків між складовими елементами бойового простору. Це обмежує здатність систем до прогнозування розвитку ситуації.

Перспективним підходом до вирішення проблеми інтерпретованості є використання когнітивних карт. *Когнітивна карта* – це структурована модель подання знань, яка відображає причинно-наслідкові зв'язки між елементами предметної області. Когнітивна карта контексту розширює цю концепцію, додаючи векторні подання станів та динаміку переходів, що допомагає інтегрувати експертні знання в архітектуру нейронних мереж [7].

Наразі постає актуальна наукова проблема – розроблення модифікованого ієрархічного трансформера з інтегрованими когнітивними картами контексту, що забезпечить ефективну обробку різнорідних даних та високий рівень інтерпретованості результатів.

Аналіз останніх досліджень і публікацій. Останніми роками активно проводяться дослідження в галузі адаптації архітектур трансформерів для обробки мультимодальної інформації. У фундаментальній роботі [5] було закладено теоретичні основи механізму самоуваги, який дав змогу досягти суттєвих результатів у обробці природної мови. Подальші дослідження дали змогу розширити використання трансформерів на інші домени, зокрема, із комп'ютерним зором [8] та мультимодальне навчання [9].

Важливий напрям досліджень пов'язаний з подоланням квадратичної обчислювальної складності $O(n^2d)$ механізму самоуваги, де n – довжина вхідної послідовності, а d – розмірність простору ознак. Це означає, що за сталої розмірності d при збільшенні кількості елементів вхідної послідовності удвічі обсяг обчислень зростає вчетверо. А це, в свою чергу, критично обмежує можливість застосування трансформерів для великих обсягів даних у реальному часі. Як свідчать результати аналізу роботи [10] існують способи для вирішення цього виклику. Автори запропонували апроксимацію, що знижує обчислювальну складність до $O(n)$. Автори використовували проєкції ключів і значень у вимір з нижчою роздільною здатністю. Але це призводить до втрат у точності моделювання далеких залежностей.

У роботі [11] автори показали можливість поєднання логічних правил із глибоким навчанням. Вони досліджували можливість інтеграції знань експертів у архітектури нейронних мереж. Але ці методи розроблялись для загальних задач. Використання їх під час математичного опису тактичної (оперативної) обстановки не було досліджено.

Чому інтерпретованість є критичною? Відповідь дають Основні положення Стратегії НАТО у сфері штучного інтелекту (Summary of the NATO Artificial Intelligence Strategy) [12]: командир не може покладатися на рішення системи з непрозорою або прихованою логікою (так званої “чорної скриньки”, англ. “black box”). Це залишається основною

перешкодою для впровадження технологій штучного інтелекту у відповідні системи збройних сил країн-членів НАТО.

Концепція когнітивних карт була вперше описана Толманом як механізм формування ментальних репрезентацій середовища для ефективного переміщення у просторі та прийняття рішень [7]. Подальший розвиток цієї ідеї пов'язаний з дослідженнями ролі гіпокампі як когнітивної карти. Так, у роботі [13] досліджено як мозок формує просторові та реляційні представлення, які в подальшому допомагають орієнтуватись в просторі з перешкодами. Ця концепція з метою використання в якості технології штучного інтелекту була формалізована Віткінгтоном та співавторами у [14]. Вони запропонували модель Tolman-Eichenbaum machine та показали, як репрезентації середовища, що структуровані за відповідними правилами, можуть покращити виконання завдань з переміщення у просторі.

У військовій сфері когнітивні карти еволюціонували за напрямом моделювання причинно-наслідкових зв'язків, зокрема, через використання нечітких когнітивних карт (fuzzy cognitive maps, FCM). Так, Коско запропонував використовувати графові структури для представлення причинної взаємодії з невизначеністю [15]. Такий підхід використаний для аналізу бойової обстановки через імітаційне моделювання, де були продемонстровані динамічні зв'язки основних факторів у складних системах. Зокрема, у роботі [16] досліджено роль теорій причинності в контексті планування бойових дій та підкреслено роль розуміння причинно-наслідкових механізмів для прийняття рішень командирами.

Насамкінець варто зазначити, що українська наукова школа має власні напрацювання в суміжних галузях. Так, у [17] вітчизняні науковці запропонували методологію системного аналізу для підтримки прийняття рішень в умовах невизначеності та багатофакторних ризиків. У роботі [18] досліджено байєсівські мережі для моделювання невизначеності в системах підтримки прийняття рішень. Розвиток цього напрямку представлено в [19], де запропоновано застосування байєсівських мереж для аналізу актуарних ризиків.

Чи можна у військовій сфері застосувати досвід, який був отриманий під час досліджень когнітивних карт? Автор вважає, що так. Проте, незважаючи на досягнення у формалізації когнітивних карт для моделювання причинно-наслідкових зв'язків [14; 15] та їх застосування для аналізу бойової обстановки [16], питання інтеграції когнітивних карт в ієрархічні трансформери для обробки різномірної інформації в інтелектуальних системах підтримки прийняття рішень залишається недостатньо дослідженою і висвітленою. Це відкриває можливості для подальших наукових досліджень.

Мета статті. Побудова модифікованого ієрархічного трансформера (архітектури глибокого навчання нейронної мережі, що побудована на механізмі самоуваги) завдяки додаванню когнітивних карт контексту для покращення процесу оброблення

різномірної інформації в інтелектуальних системах підтримки прийняття рішень військового призначення.

Виклад основного матеріалу дослідження

Для досягнення окресленої мети розглянемо спочатку обмеження стандартного механізму самоуваги, а потім запропонуємо його модифікацію. Теоретичний аналіз механізму самоуваги в класичних трансформерах [5] свідчить, що стандартна його формалізація:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

виявляє критичну проблему для використання у військових системах. Проблема полягає в тому, що механізм самоуваги розподіляє ваги на основі статистичних кореляцій у навчальних даних. Крім цього, не враховуються причинно-наслідкові зв'язки між елементами оперативної обстановки. Наприклад, вплив знищення переправи на маневрові можливості підрозділу або згасання ефекту артилерійського обстрілу в часі. Вказана проблема може бути вирішена через інтеграцію експертних знань про структуру оперативної обстановки та взаємозв'язки між її елементами безпосередньо в механізм самоуваги. Для цього пропонується використовувати ієрархічну організацію когнітивних карт контексту із застосуванням узгодження розмірностей просторів між рівнями. Це дає змогу забезпечити інтерпретованість рішень, що пропонуються системою.

Когнітивна карта контексту (рис. 1) визначається як математична структура (кортеж):

$$C = \langle S, R, E, T \rangle \quad (2)$$

де S – множина станів тактичної (оперативної) обстановки;

$R : S \times S \rightarrow [0, 1]$ – функція причинно-наслідкових зв'язків між станами тактичної (оперативної) обстановки (сила впливу стану i на стан j);

$E : S \rightarrow \mathbb{R}^d$ – вбудовування станів тактичної (оперативної) обстановки у векторний простір;

$T : S \times A \rightarrow S$ – динаміка переходів між станами тактичної (оперативної) обстановки.

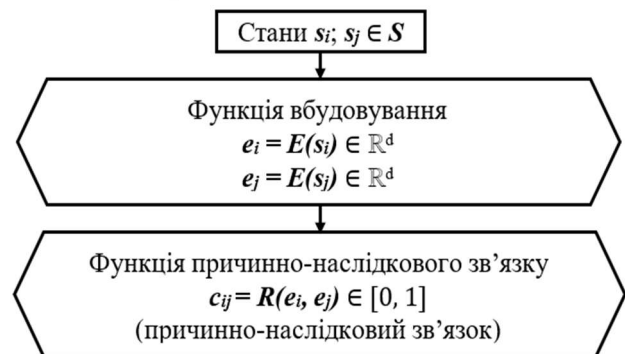


Рисунок 1 – Схема побудови причинно-наслідкових зв'язків між станами тактичної (оперативної) обстановки в когнітивній карті контексту

Для врахування різних рівнів деталізації інформації запропоновано ієрархічну організацію когнітивних карт. Визначається L рівнів ієрархії від тактичного до стратегічного, кожен з власною когнітивною картою:

$$\{C^1, C^2, \dots, C^L\}, \quad (3)$$

На думку автора, можна розрізняти шість рівнів ієрархії когнітивних карт контексту:

1. Рота.
2. Батальйон.
3. Бригада (полк).
4. Корпус (дивізія).
5. Угрупування військ (армія).
6. Оперативно-стратегічне (стратегічне) угруповання військ.

Водночас кількість рівнів ієрархії може бути змінена експериментально після проведення низки практичних заходів з опису тактичних (оперативних) ситуацій. На кожному рівні ієрархії сутності та події, що закладаються в когнітивні карти, є різними. Наприклад, на рівні роти типовими елементами когнітивної карти є позиції окремих вогневих засобів, укриття для піхотних відділень (взводів), наявність і стан боекомплекту, наявність і стан озброєння та військової техніки, маршрути висування, логістичні засоби, виявлені вогневі точки противника. На рівні батальйону елементи нижчого рівня агрегуються до бойових можливостей рот, з'являються такі нові сутності, як резерв батальйону, командно-спостережний пункт, склади батальйону, а деталізація місцевості узагальнюється до рубежів і ключових висот. На рівні бригади елементами карти стають батальйони як цілісні агреговані одиниці, артилерійські засоби, засоби протиповітряної оборони (далі – ППО) та радіоелектронної боротьби (далі – РЕБ), бригади, загальновійськовий резерв бригади, а район проведення бойових дій визначається напрямками та ключовими об'єктами, вплив на які є вирішальним для успіху операції.

Оброблення інформації на кожному конкретному рівні прийняття рішень не потребує одночасного аналізу всієї вертикалі. Командир будь-якого рівня оперує переважно трьома (чотирма) суміжними рівнями: деталізованою інформацією від підлеглих підрозділів (на 1–2 рівня нижче), власним рівнем відповідальності та контекстом від старшого командира (на 1 рівень вище). Така організація відповідає принципам військового управління, де рішення приймаються на основі доповідей знизу, оцінки обстановки та завдань старшого начальника.

Для формалізації цього принципу пропонується механізм ієрархічної уваги, який інтегрує інформацію з трьох суміжних рівнів:

$$\text{Attention}_{\text{hier}}^l = \alpha_1 W_{\uparrow}^{l-1} \text{Attention}^{l-1} + \alpha_2 \text{Attention}^l + \alpha_3 W_{\downarrow}^{l+1} \text{Attention}^{l+1}, \quad (4)$$

де α – коефіцієнти, що визначають відносну важливість кожного з трьох рівнів для конкретного завдання, а їх сума дорівнює 1.

Значення коефіцієнтів змінюються залежно від типу завдання. Наприклад, для оборонних дій зростає вага власного рівня (α_2), для наступальних – вага

нижчого рівня (α_1), для детального планування об'єднаної вогневої підтримки – вага вищого рівня (α_3);

$W_{\uparrow}^{l-1} \in \mathbb{R}^{n_l \times n_{l-1}}$ – матриця агрегації (проекція з нижчого рівня на поточний);

$W_{\downarrow}^{l+1} \in \mathbb{R}^{n_l \times n_{l+1}}$ – матриця деталізації (проекція з вищого рівня на поточний);

Проекції матриць на поточний рівень потрібні для узгодження розмірностей просторів між різними рівнями. Існуючий підхід BiFPN (bi-directional feature pyramid network – двонаправлена пірамідальна мережа) [20] використовує навчальні ваги для агрегації ознак різних масштабів. Запропонований у статті механізм ієрархічної уваги реалізує інший принцип. Він застосовує зважену агрегацію з проекційними матрицями $W_{\downarrow}$ та $W_{\uparrow}$ безпосередньо до виходів механізму уваги суміжних рівнів ієрархічних когнітивних карт. Формування когнітивних карт може здійснюватися декількома способами:

визначення експертами причинно-наслідкових зв'язків R між станами тактичної (оперативної) обстановки на основі практичного досвіду, розпорядчих і нормативних документів;

автоматичне вилучення з історичних даних операцій (бойових дій) за допомогою методів причинно-наслідкових зв'язків;

гібридний підхід, за якого автоматично вилучені зв'язки перевіряються експертами.

Для інтеграції міжрівневої інформації пропонується конкатенація когнітивних карт трьох сусідніх рівнів:

$$H^l = \text{concat}(C^{l-1}, C^l, C^{l+1}) \in \mathbb{R}^{N_{\text{total}} \times d} \quad (5)$$

де $N_{\text{total}} = N_{l-1} + N_l + N_{l+1}$ – сумарна кількість станів після конкатенації.

Для пари елементів інформації i та j матриця когнітивних впливів визначається за виразом:

$$\Phi_{ij} = R(s_i, s_j) \times \text{sim}(E(s_i), E(s_j)) \in \mathbb{R}^{N_{\text{total}} \times N_{\text{total}}} \quad (6)$$

та будується на станах конкатенованого набору:

$$S_{\text{total}}^l = S^{l-1} \cup S^l \cup S^{l+1} \quad (7)$$

де $\text{sim}(E(s_i), E(s_j))$ – функція подібності (наприклад, косинусна подібність) між векторними представленнями станів тактичної (оперативної) обстановки.

Так Φ об'єднує експертні знання про причинно-наслідкові зв'язки (R) з навченими семантичними репрезентаціями (E). Означений підхід забезпечує локальне узгодження між рівнями ієрархії когнітивних карт контексту. Отже, механізм ієрархічної уваги з узгодженням розмірностей просторів між різними рівнями реалізовує багаторівневу інтеграцію інформації. Це, в свою чергу, дає змогу моделі водночас урахувати як динамічні вагові залежності між станами тактичної (оперативної) обстановки, так і стабільні просторово-контекстуальні зв'язки між когнітивними картами різних рівнів. Для інтеграції експертних знань у механізм уваги пропонується наступна її модифікація (рис. 2) для врахування структурної інформації з когнітивної карти:

$$\text{Attention}^*(Q, K, V, \Phi) = \text{softmax}\left(\frac{QK^T + \lambda\Phi}{\sqrt{d_k}}\right)V, \quad (8)$$

де $\Phi \in S \rightarrow \mathbb{R}^{n \times n}$ – матриця когнітивних впливів (будується на основі функції вбудовування E та

причинно-наслідкових зв'язків R);

$\lambda > 0$ – гіперпараметр, який регулює баланс між даними та експертними знаннями.

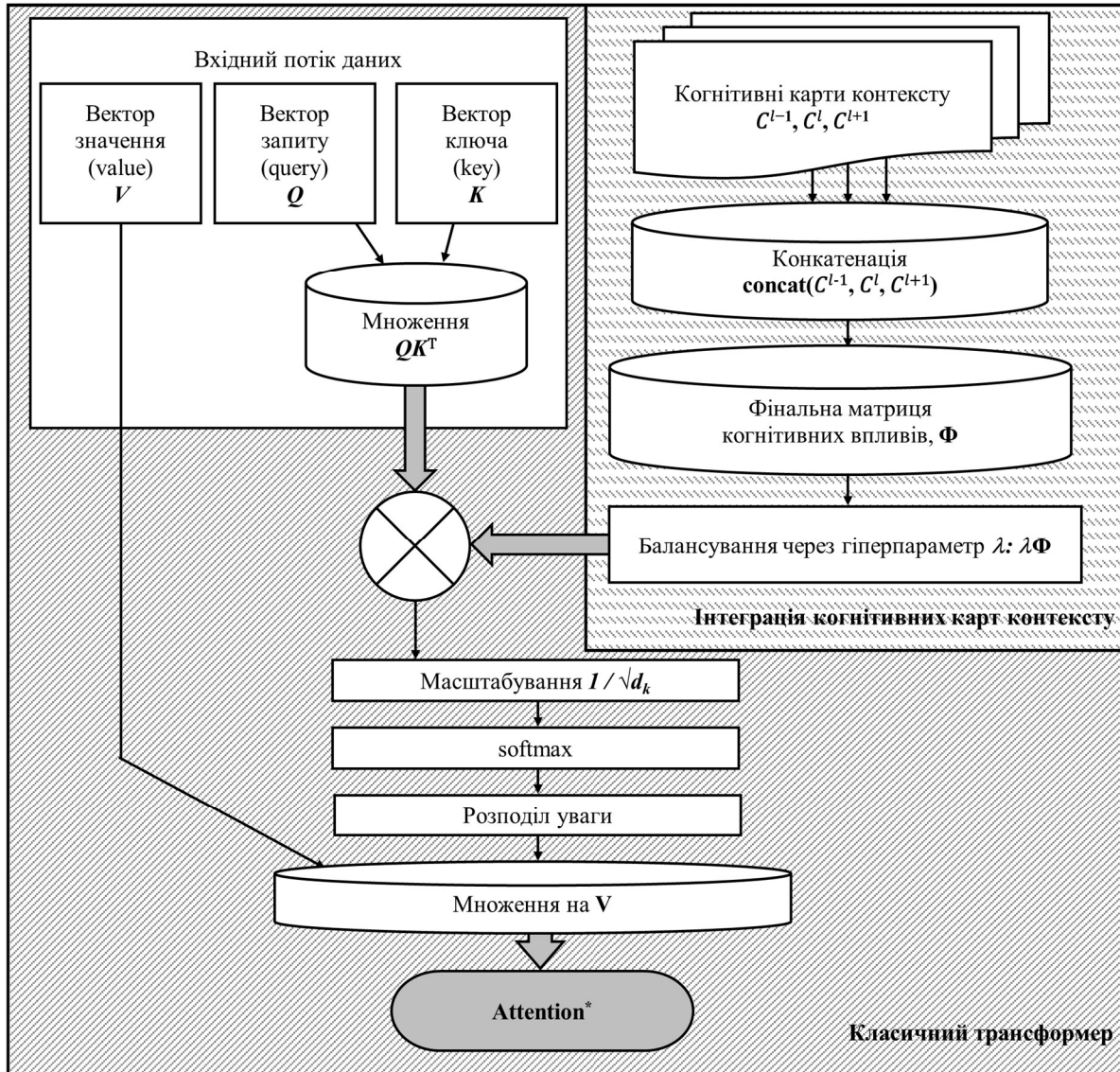


Рисунок 2 – Схема інтеграції когнітивних карт контексту до векторного простору класичного трансформера

Функція E ініціалізується випадково та навчається у процесі оптимізації моделі методом зворотного поширення помилки. Критичною проблемою військових систем є необхідність обробки асинхронних потоків даних із різними частотами оновлення. Для її вирішення запропоновано механізм темпоральної синхронізації на основі адаптивного темпорального вбудовування. Кожна модальність m отримує специфічне темпоральне кодування, що враховує її характерну динаміку, а синхронізація досягається через механізм темпоральної уваги з експоненційним згасанням впливу застарілих даних:

$$H_{\text{sync}}^m(t) = \sum_{\tau \in T_m} \alpha(t, \tau) H^m(\tau), \quad (9)$$

де: $\alpha(t, \tau) = \frac{\exp(-\gamma|t-\tau|^2)}{Z}$ – забезпечує згасання впливу застарілих даних;

T_m – множина моментів часу $\tau < t$, коли надходили дані модальності m в межах вікна спостереження,

τ – дискретні мітки часу подій;

$\gamma > 0$ – гіперпараметр швидкості згасання, специфічний для кожного типу даних (вищий для швидкозмінних даних), встановлюється експертом на основі характерного часу актуальності відповідного типу даних,

Z – нормалізуючий множник.

Попередній теоретичний аналіз свідчить, що запропонована модифікація механізму уваги (рис. 2), завдяки додаванню когнітивних карт контексту, зберігає диференційовність та сумісність зі стандартними методами оптимізації. Додавання матриці когнітивних впливів Φ до скалярного добутку QK^T не порушує властивостей градієнтного потоку, що дає змогу застосовувати стандартні оптимізатори

(Adam, AdamW). Валідація збіжності для конкретних конфігурацій когнітивних карт є предметом подальших досліджень.

Обчислювальна складність модифікованої архітектури в загальному випадку зберігає асимптотичний порядок $O(n^2d)$, властивий класичним трансформерам. Проте для військових систем управління, які характеризуються ієрархічною структурою командування та природною розрідженістю когнітивних зв'язків між віддаленими об'єктами, матриця когнітивних впливів Φ є розрідженою. Така структура матриці створює передумови для оптимізації обчислень механізму уваги, а саме розріджених тензорних подань та кластеризованих алгоритмів.

У випадку, коли середня кількість ненульових елементів у рядку матриці Φ становить $O(\log n)$, застосування розріджених тензорних подань і кластеризованих алгоритмів обчислення уваги дає змогу, за умови реалізації алгоритму розрідженої уваги [21], знизити складність до $O(n \log n \times d)$. Практична реалізація цих оптимізацій потребуватиме експериментальної перевірки на синтетичних або реальних датасетах для підтвердження збереження прийнятної якості моделювання за різних рівнів розрідженості когнітивних карт. Це буде предметом окремого дослідження. Слід окреслити обмеження запропонованого підходу:

1. Ефективність запропонованої архітектури критично залежить від якості когнітивних карт, а помилки експертів у визначенні причинно-наслідкових зв'язків можуть призводити до систематичних викривлень у прийнятті рішень.

2. Переваги розрідженої архітектури реалізуються лише за умови природної ієрархічності військових структур, що може не виконуватися в нестандартних сценаріях.

3. Інтеграція експертних знань через гіперпараметр λ потребує ретельного налаштування для кожного типу задач. Ці обмеження визначають напрями подальших досліджень.

Водночас, запропонований підхід має суттєву перевагу. Інтерпретованість рішень, які пропонуються системою, забезпечується можливістю отримання відповідних даних із когнітивних карт контексту про вплив різних факторів на елементи тактичної (оперативної) обстановки, а також зв'язки між ними.

Список бібліографічних посилань

1. Ceruti M. G., O'Brien J. J. Data management challenges and development for military information systems. *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data* (Montreal, Canada, June 4-6, 1996). New York: ACM, 1996. P. 435-445. 2. Shortland N. D., Alison L. J., Moran J. M. Military decision-making: Doctrine, rationality, and field-based approaches. *Conflict: How soldiers make impossible decisions*/ed. by P. T. Bartone, W. H. Johnston, M. V. Majchrzak. New York: Oxford University Press, 2019. P. 1-20. 3. Bengio Y., Simard P. Y., Frasconi P. Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Transactions on Neural Networks*. 1994. Vol. 5, No. 2. P. 157-166. DOI: 10.1109/72.279181. 4. Gehring J., Auli M., Grangier D., Yarats D., Dauphin Y. N. Convolutional

Висновки й перспективи подальших досліджень

У статті запропоновано модифікований ієрархічний трансформер (архітектуру глибокого навчання нейронної мережі, побудовану на механізмі самоуваги), в який інтегруються когнітивні карти контексту. Основним завданням модифікованого трансформера є обробка різномірної інформації в інтелектуальних системах підтримки прийняття рішень військового призначення.

Запропонована формальна математична модель когнітивної карти контексту (2) дає змогу явно описувати причинно-наслідкові зв'язки між елементами тактичної (оперативної) обстановки та створює основу для інтерпретованості рішень, які пропонуються системою. Модифікований механізм самоуваги з інтеграцією матриці когнітивних впливів дає змогу поєднати статистичне навчання на даних із експертними знаннями про структуру бойового простору. Ієрархічна організація когнітивних карт із механізмом узгодження розмірностей між рівнями відповідає принципам військового управління та забезпечує обробку інформації на різних рівнях деталізації – від тактичного до стратегічного. Механізм темпоральної синхронізації вирішує проблему асинхронності інформаційних потоків, забезпечуючи коректне врахування актуальності даних через експоненційне згасання впливу застарілої інформації.

Перспективи подальших досліджень передбачають:

експериментальну валідацію запропонованої архітектури на синтетичних датасетах тактичних (оперативних) епізодів;

розроблення методів автоматизованого вилучення когнітивних карт контексту з історичних даних бойових дій;

дослідження впливу різних рівнів розрідженості матриці когнітивних впливів на точність прогнозування;

розроблення адаптивних механізмів визначення гіперпараметрів λ та γ залежно від рівня невизначеності та повноти вхідних даних.

Sequence to Sequence Learning. *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*. Sydney, Australia, 2017. Vol. 70. P. 1243-1252. DOI: 10.48550/arXiv.1705.03122. 5. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017. Vol. 30. P. 5998-6008. DOI: 10.48550/arXiv.1706.03762. 6. Alam F. S., Al-Mamun M. A. Transformers and large language models for efficient intrusion detection systems: A comprehensive survey. *Information Fusion*. 2025. Vol. 113. 102567. 7. Tolman E. C. Cognitive maps in rats and men. *Psychological Review*. 1948. Vol. 55. № 4. P. 189-208. URL: <https://personal.utdallas.edu/~tres/spatial/tolman.pdf> (accessed:

- 15 September 2025). 8. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houshy N. An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. 2021. DOI: 10.48550/arXiv.2010.11929. 9. Radford A., Kim J. W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. Learning transferable visual models from natural language supervision. *International Conference on Machine Learning*. 2021. Vol. 139. P. 8748–8763. DOI: <https://doi.org/10.48550/arXiv.2103.00020>. 10. Wang S., Li B., Khabsa M., Fang H., Ma H. Linformer: Self-Attention with Linear Complexity. 2020. DOI: 10.48550/arXiv.2006.04768. 11. Garcez A. d'A., Lamb L. C. Neurosymbolic AI: The 3rd wave. *Artificial Intelligence Review*. 2023. Vol. 56. № 11. P. 12387–12406. DOI: 10.1007/s10462-023-10448-w. 12. Summary of the NATO Artificial Intelligence Strategy. NATO. 2021. URL: <https://nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy> (accessed: 15 September 2025). 13. O'Keefe J., Nadel L. The Hippocampus as a Cognitive Map. Oxford: Clarendon Press, 1978. 570 p. URL: <https://www.cmor-faculty.rice.edu/~cox/neuro/HCM/Complete.pdf> (accessed: 15 September 2025). 14. Whittington J., Muller T., Mark S., Chen G., Barry C., Burgess N., Behrens T. The Tolman-Eichenbaum machine: Unifying space and relational memory through generalisation in the hippocampal formation. *Cell*. 2020. Vol. 183. № 5. P. 1249–1263. DOI: 10.1016/j.cell.2020.10.024. 15. Kosko B. Fuzzy cognitive maps. *International Journal of Man-Machine Studies*. 1986. Vol. 24. P. 65–75. DOI: 10.1016/S0020-7373(86)80040-2. 16. Diehl P. A Study of Causality in Military Planning: Master's thesis. Fort Leavenworth, KS: U.S. Army Command and General Staff College, 2012. URL: <https://apps.dtic.mil/sti/tr/pdf/ADA563134.pdf> (accessed: 15 September 2025). 17. Zgurovsky M., Pankratova N. System Analysis & Intelligent Computing: Theory and Applications. Cham: Springer, 2022. 419 p. DOI: 10.1007/978-3-030-94910-5. 18. Згуровський М. З., Білюк П. І., Терент'єв О. М., Просянкін-Жарова Т. І. Байєсівські мережі в системах підтримки прийняття рішень. Київ: ТОВ «ВП «Едельвейс»», 2015. 300 с. 19. Panibratov R., Bidyuk P. Applying Bayesian networks in analysis of actuarial risks. *Problems of Control and Informatics*. 2025. № 3. DOI: 10.34229/1028-0979-2025-3-3. 20. Tan M., Pang R., Le Q. EfficientDet: Scalable and Efficient Object Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020. P. 10781–10790. 21. Child R., Gray S., Radford A., Sutskever I. Generating Long Sequences with Sparse Transformers. 2019. DOI: 10.48550/arXiv.1904.10509.

MODIFIED HIERARCHICAL TRANSFORMER IN INTELLIGENT DECISION SUPPORT SYSTEMS FOR MILITARY APPLICATIONS

SINEGLAZOV Viktor, Doctor of Technical Sciences, Professor, State University «Kyiv Aviation Institute», Kyiv, Ukraine, <https://orcid.org/0000-0002-3297-9060>

CHELKOVAN Viktor, Central Research Institute of the Armed Forces of Ukraine, Kyiv, Ukraine, <https://orcid.org/0009-0001-4522-6260>

Formulation of the problem in general. Today, combat actions are characterised by the need to process significant volumes of heterogeneous data, including intelligence, meteorological, and geospatial data. They have different update dynamics and presentation formats. Traditional approaches to integrating such data have several problems. Cause-and-effect relationships between elements of the combat situation are not taken into account, solutions are not interpretable enough, and the quadratic computational complexity of directly applying classical transformers is high.

Literature review. Fundamental works on transformer architectures and their application to multimodal information processing have been analysed. Studies on reducing the computational complexity of the self-attention mechanism and integrating expert knowledge into neural architectures in the context of neuromyotonic artificial intelligence have been reviewed. The contributions of Ukrainian scientists to the methodology of systems analysis for decision support under uncertainty, as well as the application of Bayesian networks for modelling uncertainty, have been investigated. The evolution of the cognitive map concept, from its neuroscientific foundations to the application of fuzzy cognitive maps in the military domain for modelling causal relationships, has been analysed. It has been established that the integration of cognitive maps into transformer architectures for processing heterogeneous information in intelligent decision support systems for military purposes, particularly with an emphasis on causal relationships in the battlespace, remains insufficiently developed and requires theoretical justification.

Research results. Theoretical approaches to modifying a hierarchical transformer with integrated cognitive context maps have been formulated. A formal mathematical model of the cognitive context map is proposed as a structure that describes the set of states of the tactical (operational) situation, causal relationships between them, their vector representations, and the dynamics of transitions. A modified self-attention mechanism that integrates a matrix of cognitive influences is developed. The hierarchical organisation of cognitive context maps with a mechanism for aligning the dimensions of spaces between levels is justified. A mechanism for temporal synchronisation of processing asynchronous data streams, based on the adaptive decay of the influence of outdated information, is proposed.

Research novelty. A novel modification of the transformer self-attention mechanism is proposed, achieved through integration of formalised cognitive context maps. This approach ensures decision interpretability via explicit representation of causal relationships among combat situation elements. It also enables uniform processing of heterogeneous information: encompassing geospatial object coordinates, temporal dependencies with influence decay effects, and discrete events affecting objects over time. To support the hierarchical organisation of cognitive maps, a dimensionality alignment mechanism is proposed across levels, ensuring the aggregation of spatial-contextual information.

The theoretical and practical significance of this article lies in establishing a mathematical and algorithmic foundation for developing intelligent decision support systems in military applications. A key function of such systems is real-time processing of heterogeneous information streams from various sources. The proposed transformer provides a high level of decision interpretability through a tactical (operational) situation decomposition mechanism. This modified architecture with integrated cognitive maps enables commanders to understand the causal relationships between situation elements and system-generated decisions.

Conclusions and future work. The proposed modified transformer with integrated cognitive maps enables the effective processing of heterogeneous information through the modification of the self-attention mechanism, hierarchical organisation of cognitive maps, and temporal synchronisation of asynchronous information streams. Expert knowledge integration via the cognitive influence matrix ensures decision interpretability and enables multi-criteria decomposition. Future research directions include the experimental validation of the proposed transformer using real or synthetic datasets of tactical (operational) combat episodes, as well as the development of methods for automated extraction of cognitive context maps from historical data.

Keywords: transformers, cognitive maps, decision support systems, multimodal data processing, self-attention mechanism, artificial intelligence, hierarchical data integration, temporal synchronisation.

References

1. Ceruti, M. and O'Brien, J., (1996). Data Management Challenges and Development for Military Information Systems. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data. Montreal, Canada, 4-6 June 1996. New York: ACM, 435-445.
2. Shortland, N., Alison, L. and Moran, J., (2019). Military decision-making: Doctrine, rationality, and field-based approaches. In: Bartone, P., Johnston, W. and Majchrzak, M., eds. Conflict: How soldiers make impossible decisions. New York: Oxford University Press, 1-20.
3. Bengio, Y., Simard, P. Y., & Frasconi, P., (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157-166. DOI: 10.1109/72.279181.
4. Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N., (2017). Convolutional sequence-to-sequence learning. In D. Precup & Y. W. Teh (Eds.), Proceedings of the 34th International Conference on Machine Learning (Vol. 70, pp. 1243-1252). PMLR. Proceedings of Machine Learning Research DOI: 10.48550/arXiv.1705.03122.
5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L. and Polosukhin, I., (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008. DOI: 10.48550/arXiv.1706.03762.
6. Alam, F. and Al-Mamun, M., (2025). Transformers and Large Language Models for Efficient Intrusion Detection Systems: A Comprehensive Survey. Information Fusion, 113, 102567.
7. Tolman, E. C., (1948). Cognitive maps in rats and men. Psychological Review, 55(4), 189-208. Available at: <https://personal.utdallas.edu/~tres/spatial/tolman.pdf> [Accessed: 15 September 2025].
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Housby, N., (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR). DOI: 10.48550/arXiv.2010.11929.
9. Radford, A., Kim, J., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. and Sutskever, I., (2021). Learning transferable visual models from natural language supervision. International Conference on Machine Learning, 139, 8748-8763. DOI: 10.48550/arXiv.2103.00020.
10. Wang, S., Li, B., Khabsa, M., Fang, H. and Ma, H., (2020). Linformer: Self-Attention with Linear Complexity [online]. Available at: <https://arxiv.org/abs/2006.04768> [Accessed: 15 September 2025].
11. Garcez, A. d'A. and Lamb, L. C., (2023). Neurosymbolic AI: The 3rd wave. Artificial Intelligence Review, 56(11), 12387-12406. DOI: 10.1007/s10462-023-10448-w.
12. Summary of the NATO Artificial Intelligence Strategy (2021). Brussels: NATO Headquarters. [online]. Available at: <https://nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy> [Accessed: 15 September 2025].
13. O'Keefe, J. and Nadel, L., (1978). The Hippocampus as a Cognitive Map. Oxford: Clarendon Press. Available at: <https://www.cmorf.aculty.rice.edu/~cox/neuro/HCMComplete.pdf> [Accessed: 15 September 2025].
14. Whittington, J., Muller, T., Mark, S., Chen, G., Barry, C., Burgess, N. and Behrens, T., (2020). The Tolman-Eichenbaum machine: Unifying space and relational memory through generalisation in the hippocampal formation. Cell, 183(5), 1249-1263. DOI: 10.1016/j.cell.2020.10.024.
15. Kosko, B., (1986). Fuzzy cognitive maps. International Journal of Man-Machine Studies, 24, 65-75. DOI: 10.1016/S0020-7373(86)80040-2.
16. Diehl, P., (2012). A Study of Causality in Military Planning. Master's thesis. Fort Leavenworth, KS: U.S. Army Command and General Staff College. Available at: <https://apps.dtic.mil/sti/tr/pdf/ADA563134.pdf> [Accessed: 15 September 2025].
17. Zgurovsky, M. and Pankratova, N. eds., (2022). System Analysis & Intelligent Computing: Theory and Applications. Cham: Springer. DOI: 10.1007/978-3-030-94910-5.
18. Zgurovsky, M. Z., Bidyuk, P. I., Terentiev, O. M. and Prosyankina-Zharova, T. I., (2015). Bayesian networks in decision support systems. Kyiv: Edelweiss.
19. Panibratov, R. and Bidyuk, P., (2025). Applying Bayesian networks in the analysis of actuarial risks. Problems of Control and Informatics, 3. DOI: 10.34229/1028-0979-2025-3-3.
20. Tan, M., Pang, R. and Le, Q., (2020). EfficientDet: Scalable and Efficient Object Detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 10781-10790.
21. Child, R., Gray, S., Radford, A. and Sutskever, I., (2019). Generating Long Sequences with Sparse Transformers [online]. Available at: <https://arxiv.org/abs/1904.10509> [Accessed: 15 September 2025].

Рукопис надійшов до редакції 16.10.2025
 Рукопис прийнято до друку після рецензування 09.12.2025
 Дата публікації 30.12.2025