

МЕТОД ПАРАМЕТРИЧНОЇ ОБРОБКИ ІНФОРМАЦІЇ В ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ СИСТЕМАХ

Розглянуто основні параметри, що визначають важливість першочергової обробки вхідних електронних документів в інформаційно-аналітичних системах. Проаналізовано можливість використання нейронних мереж для вирішення задач визначення важливості вхідних документів. Як об'єкт, на якому було проведено відповідні дослідження, вибрано систему підтримки прийняття рішення щодо першочерговості розглядання вхідних документів. З цією метою проводиться аналіз статистичних, функціональних та технічних показників, що характеризують важливість документів. Обґрунтовано можливість застосування нечіткого гібридного класифікатора для визначення важливості вхідних документів з метою обґрунтування послідовності їх обробки в інформаційно-аналітичній системі. Запропоновано використовувати 4-прошаркову нейро-нечітку мережу для підвищення ефективності розгляду вхідних документів.

Ключові слова: вхідні електронні документи, нейронні мережі, важливість вхідних документів, нечіткий гібридний класифікатор.

Вступ. Вирішення задачі оцінки важливості вхідних документів (ВД) в інформаційно-аналітичній системі (ІАС) обробки інформації значно загострилося в сучасних умовах, коли різко збільшились обсяги постачання вхідних документів, а часові витрати, необхідні для обробки інформації, залишаються незмінними. Саме тому задачі визначення важливості ВД з метою першочергової обробки найбільш важливих – займають провідне місце в сучасних ІАС.

На даний час процес оцінки важливості ВД стикається зі значними розбіжностями у формуванні і інтерпретації комплексу інформативних параметрів та способів їх обробки [1, 2]. Крім того, аналіз комплексу таких інформативних параметрів показує наявність великої кількості різнорідних значень, що вимірюються в порядковій шкалі, шкалах інтервалів, відношень і абсолютній шкалі. Переважна більшість існуючих методів обробки експериментально отриманої інформації не пристосована для врахування різнорідності (різношкальності) значень виміряних параметрів.

Метою даної статті є обґрунтування методу параметричної обробки та визначення послідовності розгляду ВД з урахуванням їх важливості.

Виклад основного матеріалу дослідження. На даний час в ІАС використовується послідовний засіб перегляду кожного документа без урахування його важливості. Проте для зберігання в системі повних текстів документів потрібні великі витрати часу на перенесення цих текстів до пам'яті ЕОМ. Крім того, для зберігання інформації у повному обсязі необхідно багато пам'яті, а сам процес розгляду ВД триватиме досить довго. Тому в ІАС доцільно забезпечити першочерговий розгляд найбільш важливих ВД.

Важливість ВД неможливо охарактеризувати як просту змінну в протіканні окремих функцій чи процесів. Це складна системна величина, що характеризує зміст та терміновість відпрацювання ВД, пріоритет адресата, який надіслав ВД. Під "системою" при цьому розуміють сукупність взаємодіючих між собою елементарних структур чи процесів, поєднаних в одне ціле рішенням спільної задачі, яка не може бути виконана жодним з компонентів системи [3, 4, 5].

Таким чином, важливість ВД змінюється з часом і не піддається опису за допомогою набору фіксованих параметрів (наприклад за ключовими словами) через наступні причини:

можливе різне написання одних і тих самих слів;

серед ключових слів є дуже багато синонімів та омонімів;

ключові слова не визначають родово-видові відношення між словами.

Роздільна обробка виміряних показників важливості ВД і критеріїв її оцінювання не завжди дозволяє дати однозначну оцінку [6]. Багаторівневість важливості ВД в ІАС задає різні напрямки для її визначення.

Один з таких напрямків полягає в пошуку усталених конфігурацій відповідей для різних поведінкових ситуацій, що є базою для виділення інтегральних показників для оцінки важливості ВД.

У загальному випадку під інтегральними методами розуміють методи, що засновані на об'єднанні декількох методик визначення окремих показників або невеликої сукупності однорідних показників для виведення на їх основі умовної („штучної”) оцінки важливості ВД.

Велика кількість факторів, від котрих залежить важливість ВД, а також різноманітність функцій, в яких проявляється його специфічність, є основною складністю у вирішенні задач оцінки ефективності ІАС. Вирішити ці задачі можна лише шляхом використання інтегральних методів визначення важливості ВД.

Проблеми невизначеності і багатофакторності виникають як в середині кожної складової ВД, так і при згортці сукупності оцінок в інтегральний (узагальнений) показник важливості ВД ($K_{\text{од}}$). Так, наприклад, за результатами аналізу джерел [1-5] для оцінки важливості ВД можна використати наступні показники (параметри):

а) статистичні показники \vec{X}^{Φ} :

– щільність ключових слів у назві та в анотації ВД, що характеризує важливі та найважливіші ВД (ЩНА), шкала вимірювання – відношення;

– щільність ключових слів у тексті, що характеризує найважливіші ВД (ЩНД), шкала вимірювання – відношення;

– щільність ключових слів у тексті, що характеризує важливі ВД (ЩВД), шкала вимірювання – відношення;

– щільність ключових слів у тексті, що характеризує не важливі ВД (ДВД), шкала вимірювання – відношення;

б) функціональні показники \vec{Y}^{Π} :

– виходячи із щільності ключових слів – клас (КВД), до якого належить ВД, шкала вимірювання – порядкова;

– мотивація (М) необхідності першочергової обробки ВД за часом, шкала вимірювання – порядкова;

в) технічні показники \vec{Z}^E :

– важливість абонента (ВА), який надіслав ВД, шкала вимірювання – порядкова.

З урахуванням вказаних параметрів ВД блок-схему розрахунку $K_{\text{од}}$ можна представити у вигляді, зображеному на рис. 1.

На рис. 1 через $\vec{\rho}_{\text{ЩНД} / \text{ЩВД}}$ та $\vec{\rho}_{\text{ДВД} / \text{М}}$ враховуються експертні висновки про існування залежностей між окремими показниками однієї або кількох груп показників.

Для вирішення задачі класифікації, тобто віднесення ВД, що характеризується набором показників, до важливих або найбільш важливих, пропонується застосувати так званий нечіткий гібридний класифікатор [5-6]. Такий класифікатор є системою, що об'єднує в структурному і функціональному відношеннях принципи нейронних мережних моделей і нечітку логіку обробки даних відповідно.

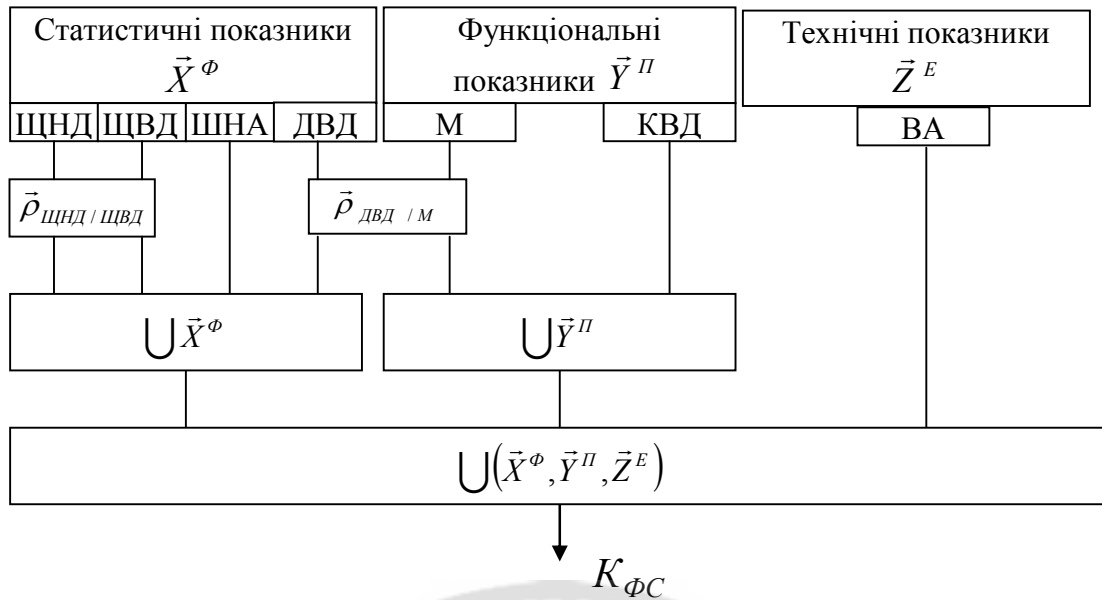


Рис. 1. Блок-схема розрахунку узагальненого показника важливості ВД

Поставлену задачу будемо вирішувати за допомогою 4-прошаркової нейро-нечіткої мережі, структурна схема якої наведена на рис. 2.

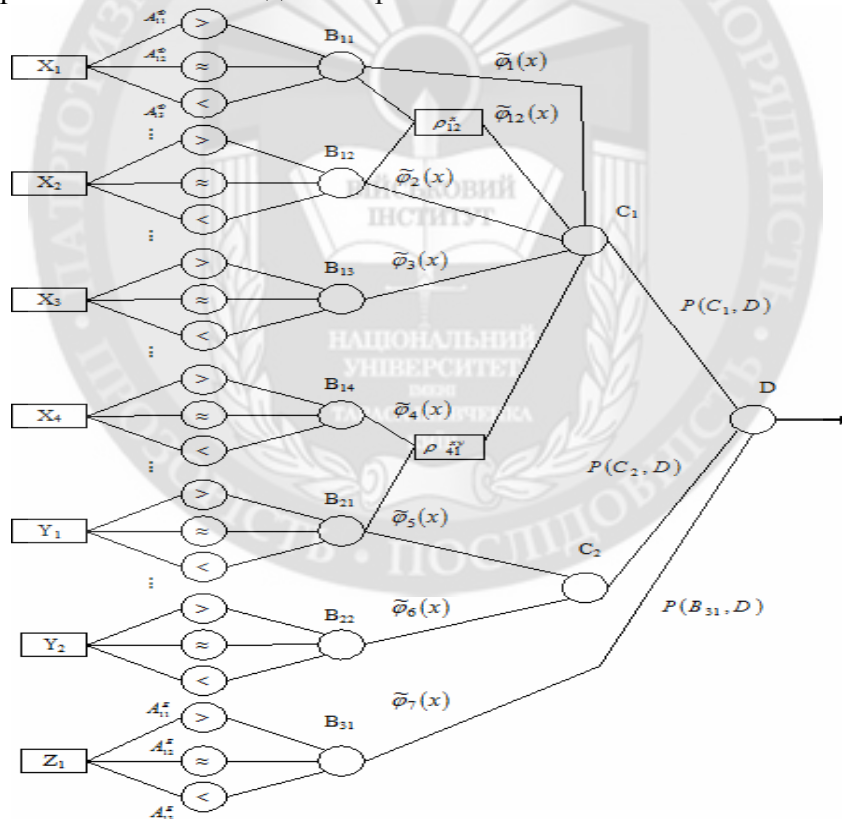


Рис. 2. Структурна схема прошаркової нейро-нечіткої мережі.

Перший прошарок A мережі створює на виході ступінь належності як міру відповідності вимірних показників ВД $\{X, Y, Z\}$ заданим вимогам.

Запропонований варіант мережі розрахований на 3-рівневе нечітке оцінювання: „неважливе” („н” або $<$), „важливе” („в” або \approx), „найважливіше” („нв” або $>$). Типовий приклад функцій розподілення для вказаних лінгвістичних оцінок наведено на рис. 3.

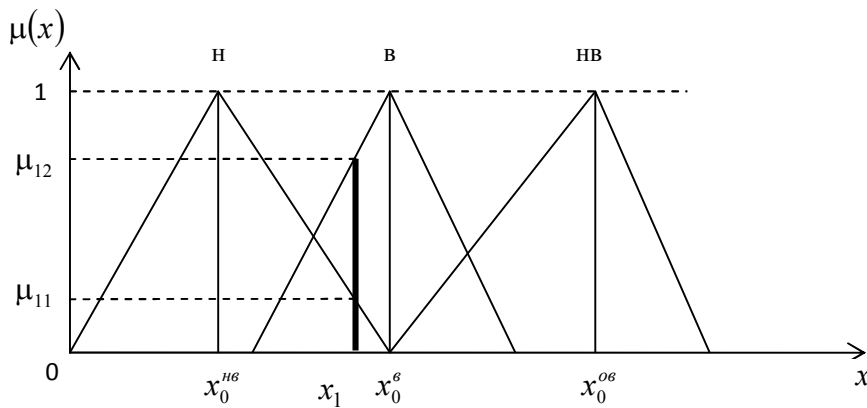


Рис. 3. Приклад функцій розподілення лінгвістичних оцінок

При необхідності роздільна здатність такого класифікатора може бути підвищена, що тягне за собою збільшення числа нейронів у прошарку A , але це не вплине на інші прошарки мережі та алгоритм її функціонування.

Другий прошарок B є об'єднуючим по кожному конкретному показнику $x_i (y_j; z_k)$ і потрібен для того, щоб врахувати можливість попадання i -го признаку одночасно в дві класифікаційні групи (як правило, з різним ступенем належності). Наприклад (див. рис. 2):

$$B_{11} = \tilde{\varphi}_1(x) = \{A_{11}^\phi | \mu_{11}(x_1); A_{12}^\phi | \mu_{12}(x_1)\},$$

де μ_{11} – міра відповідності результату вимірювання ЩНД вимозі „нижче норми”, μ_{12} – міра відповідності результату вимірювання ЩНД вимозі „норма”.

Третій прошарок C призначений для об'єднання нечітких оцінок всередині кожної групи показників: \vec{X}^ϕ , \vec{Y}^π і \vec{Z}^E . Особливістю зв'язків між 3 і 4-м прошарками є наявність всередині групових ρ_{ij}^x і міжгрупових ρ_{ij}^{xy} допоміжних функціональних елементів, що відображають факт функціональної залежності відповідних (i, j) показників. Введення допоміжних функціональних елементів дозволяє підвищити гнучкість і надійність роботи класифікатора в умовах можливої неповноти вимірювань з причини часових або технічних проблем у штатному режимі.

Нейрони прошарку C – це стандартні нейрони, виходи яких формуються з використанням активаційних функцій сигмоїдального типу і трактуються як ступені належності (міри відповідності) статистичних (функціональних, технічних) показників заданим вимогам визначення важливості ВД.

Четвертий прошарок D представлений єдиним нейроном, входами якого є зважені значення мір відповідності ВД по кожній групі показників, а виходом – міра відповідності щодо першочерговості розглядання ВД.

Ваги зв'язків $P(C_i, D)$ між третім і четвертим прошарками визначаються експертами заздалегідь в залежності від конкретного класу ВД і характеризують важливість тієї чи іншої групи показників.

Запропонована нейро-нечітка мережа може бути класифікована як синхронна багатопрешаркова гетерогенна мережа з локальними зв'язками, але без зворотних зв'язків. Останнє дозволяє зняти питання про динамічну врівноваженість нейромережі, що є важливою перевагою наведеної структури.

Подальші дослідження доцільно спрямувати на підвищення достовірності функціонування нейро-нечіткого класифікатора за рахунок об'єктивного обґрунтування ваги міжпрешаркових зв'язків $P(C_i, D)$ та кореляційних залежностей ρ_{ij}^x і ρ_{ij}^{xy} .

Висновок. Запропоноване рішення для оцінки важливості вхідних електронних документів із застосуванням нечіткого гібридного класифікатора доцільно покласти в основу

системи підтримки прийняття рішень щодо першочерговості розглядання ВД в інформаційно-аналітичних системах.

ЛІТЕРАТУРА:

1. Интеллектуальные системы поддержки принятия решений: теория, синтез, эффективность / В.О. Тарасов, Б.М. Герасимов, И.О. Левин, В.О. Корнищук. – К.: МАКНС, 2007. – 255 с.
2. Герасимов Б.М. Человеко-машинные системы принятия решений с элементами искусственного интеллекта / Герасимов Б.М., Тарасов В.А., Токарев И.В. – К.: Наукова думка, 1993. – 184 с.
3. Герасимов Б.М. Нечеткие множества в задачах проектирования, управления и обработки информации / Герасимов Б.М., Грабовский Г.Г., Рюмшин Н.А. – К.: Техніка, 2002. – 140 с.
4. Герасимов Б.М. Система поддержки принятия решений в АСУ реального времени / Герасимов Б.М., Глуцкий В.М., Рабчун А.А. // Искусственный интеллект. – № 3.– 2000.– С. 39-47.
5. Круглов В.В. Нечеткая логика и искусственные нейронные сети / Круглов В.В., Длин М.И., Голунов Р.Ю. – М.: Физматлит, 2001. – 225 с.
6. Мельник Ю.В. Застосування взаємодіючих нейромереж в задачах визначення готовності льотних екіпажів / Ю.В. Мельник, О.Ю. Чуніхін // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – 2007.– Вип. 7. – С. 123-128.

Рецензент: д.т.н., проф. Ленков С.В., начальник науково-дослідного центру Військового інституту Київського національного університету імені Тараса Шевченка

Шворов А.С.

МЕТОД ПАРАМЕТРИЧЕСКОЙ ОБРАБОТКИ ИНФОРМАЦИИ В ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМАХ

Рассмотрены основные параметры, которые определяют важность первоочередной обработки входных электронных документов в информационно-аналитических системах. Проанализирована возможность использования нейронных сетей для решения задач определения важности входных документов. Как объект, на котором были проведены соответствующие исследования, выбрана система поддержки принятия решения относительно первоочередного рассмотрения входных документов. С этой целью проводится анализ статистических, функциональных и технических показателей, которые характеризуют важность документов. Обоснована возможность применения нечеткого гибридного классификатора для определения важности входных документов с целью обоснования последовательности их обработки в информационно-аналитической системе. Предложено использовать 4-слойную нейро-нечеткую сеть для повышения эффективности рассмотрения входных документов.

Ключевые слова: входные электронные документы, нейронные сети, важность входных документов, нечеткий гибридный классификатор.

Shvorov A.

METHOD PARAMETRIC INFORMATION PROCESSING IN INFORMATION-ANALYTICAL SYSTEM

Basic parameters that determine importance of primary treatment of entrance electronic documents in the research and information systems are considered. Possibility of the use of neural networks is analyzed for the decision of tasks of determination of importance of entrance documents. As an object on that corresponding studies were undertaken, the system of support of decision-making of relatively primary consideration of entrance documents is chosen. The analysis of statistical, functional and technical indexes that characterize importance of documents is conducted to that end. Possibility of application of unclear hybrid classifier is reasonable for determination of importance of entrance documents with the purpose of ground of sequence of their treatment in the research and information system. It is offered to use a 4-layer neuro-indistinct network for increase of efficiency of consideration of entrance documents.

Keywords: entrance electronic documents, neural networks, importance of entrance documents, unclear hybrid classifier.