

УДК 004.67+519.6

О. М. Терентьев, Т. І. Просьянкіна-Жарова, В. В. Савастьянов

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, 03056 Київ, Україна

Використання засобів текстової аналітики як інструменту оптимізації підтримки прийняття рішень у задачах розробки планів соціально-економічного розвитку України

Розглянуто питання застосування інструментів SAS Text Mining у задачах підтримки прийняття рішень. Інструменти текстової аналітики застосовують для автоматизації попереднього аналізу проблеми, визначення факторів впливу зовнішнього середовища у задачі дослідження розвитку сільського господарства, виходячи із необхідності переорієнтації ринків збуту та забезпечення продовольчої безпеки країни. Проаналізовано основні проблеми аналізу великих обсягів неструктурованої текстової інформації у задачах багатокритеріального аналізу альтернатив у системах підтримки прийняття рішень. За допомогою контент-аналізу та аналізу емоційного забарвлення інтернет-публікацій можна значно покращити якість прийнятих рішень завдяки опрацюванню більшої кількості експертних думок та уявлень про можливий перебіг подій і формалізувати задачу цільової установки тощо. Розроблено загальні рекомендації щодо застосування інструментів SAS Text Mining у роботі інформаційно-аналітичних центрів органів державного управління розвитком агропромислового виробництва.

Ключові слова: слабкоструктуровані задачі, текстова інформація, багатокритеріальний аналіз альтернатив, система підтримки прийняття рішень, інтернет-ресурси, сільське господарство, text mining.

Вступ

Управління розвитком соціально-економічних систем в умовах суспільно-політичних перетворень за необхідності забезпечення сталого розвитку та продовольчої безпеки країни, передбачення загроз національній безпеці, забезпечення контрольованості суспільно-політичної ситуації у країні, як відзначається у програмі НАТО «Наука заради миру та безпеки» [1] є пріоритетними для багатьох країн, що розвиваються, в тому числі і для України.

© О. М. Терентьев, Т. І. Просьянкіна-Жарова, В. В. Савастьянов

Питання необхідності оптимізації інформаційно-аналітичної підтримки у сфері управління національною економікою України є одним із найбільш актуальних і розглядається у роботах багатьох вітчизняних учених, зокрема О. Дегтяря [2], Н. Дяченка [3], В. Алексєєва і В. Терещенка [4] та ін. Як відзначають М.З. Швиденко [5], Ю.О. Лупенко, В.М. Жук, Б.В. Мельничук і В.М. Метелиця [6], В.В. Харченко і В.Р. Оніщук [7] та інші, надзвичайно важливо сьогодні створювати відповідні галузеві (або загальнодержавні) інформаційно-аналітичні системи, основним елементом яких є обробка інформації, що зберігається на розподілених ресурсах.

Необхідність такої обробки зумовлена як обґрунтуванням цільових настанов, так і дослідженням альтернатив у задачах багатокритеріального вибору, методи вирішення яких розглядаються у роботах Т. Сааті [8], М.З. Згуровського [9], Н.Д. Панкратової і Н.І. Недашківської [10], О.І. Ларічева [11], В.А. Шакірова та П.С. Панкратьєва [12], З.К. Авдєєвої, С.В. Ковриги [13] та інших.

Як відзначають А.А. Барсегян та ін. [14], М.В. Берри [15], С. Аггарвала та С. Жао [16], Д. Прадо та Е. Фернеда [17], Н.Д. Панкратова та І.О. Савченко [18], виявити дійсно значимі альтернативи, дослідити закономірності зв'язків між подіями, виконати категоризацію висновків експертів і представити цю інформацію у вигляді, оптимальному для особи, що приймає рішення, можна за допомогою інструментів Text Mining. Однак, не зважаючи на те, що текстова аналітика визнана важливим та ефективним інструментом підтримки прийняття рішень, питання її застосування у інформаційно-аналітичному забезпеченні вирішення задач багатокритеріального вибору залишається недостатньо опрацьованим.

Формулювання мети

Мета дослідження полягає в обґрунтуванні вибору оптимальних інструментів, методик, підходів і способів їхнього застосування для вирішення задач системних інформаційно-аналітичних досліджень і підвищення результативності таких досліджень у широкому спектрі задач, зокрема, у вирішенні задач забезпечення продовольчої безпеки країни шляхом удосконалення існуючих методів, що пов'язані із передбаченням в умовах невизначеностей і слабкої формалізації задач.

Виклад основного матеріалу

Після подій 2013–2014 рр., національна економіка України продовжує функціонувати в умовах суспільно-політичної та фінансової нестабільності, що супроводжується радикальними змінами в галузевій структурі, стрімкою переорієнтацією зовнішньоекономічних зв'язків тощо. За таких умов потребує перегляду та доповнення існуюча методика макроекономічного прогнозування та планування, адже, як відзначає М.З. Згуровський [9, с. 58], сьогодні актуальною є методологія, яка «дає можливість уявити майбутнє, яке не може бути інтерпретоване як звичайне продовження минулого, оскільки це майбутнє набуває принципово нового змісту, форм і структур».

Погоджуючись із думкою З.К. Авдєєвої та ін. [13], що однією з головних задач, які треба розв'язувати при розробці моделей і методів прийняття рішень у слабкоструктурованих ситуаціях, є задача формалізації нечітких уявлень, слід від-

значити, що частково, дану проблему можна вирішити шляхом застосування сучасних методів аналізу структурованих та неструктурованих даних із текстових масивів Інтернету та офіційних статистичних даних.

Враховуючи, що обсяги інформації, яка відображає уявлення експертів про соціальні, політичні та економічні процеси, стрімко накопичуються у відкритих інтернет-джерелах, актуальною стає задача обробки таких текстових ресурсів як блоги, портали новин, статті в електронних виданнях, обговорення тощо, з метою цілеспрямованого видобування знань. Традиційно, для цих цілей застосовується розвідувальний пошук, коли аналізувалися документи з різних джерел, у тому числі і з інтернет-ресурсів, у результаті чого, дослідник одержував набір відомостей за темою пошукового запиту та виявляв підтеми досліджуваної проблеми, що в подальшому потребувало обробки отриманих результатів експертами, підвищуючи при цьому суб'єктивність прийняття управлінських рішень.

Проблема оптимізації інструментарію прийняття управлінських рішень щодо обґрунтування сценаріїв розвитку національної економіки, її окремих секторів і галузей нерозривно пов'язана з формуванням об'єктивного уявлення експертів щодо поточного стану національної економіки, виявлення потенційних проблем, які можуть виникати за даного розвитку подій. Складність вирішення цієї задачі значною мірою зумовлена тим, що для побудови прогнозів за допомогою математичних та економетричних моделей часто неможливо сформувати достатньо довгі часові ряди співставних статистичних показників соціально-економічного розвитку України. Тому все більше дослідників віддають перевагу використанню когнітивних, причинно-наслідкових моделей у поєднанні з математичними та економетричними моделями, що значно підвищує вимоги до інформаційного забезпечення процесу підтримки прийняття рішень. Інформаційною основою для побудови таких моделей, враховуючи постійне зростання обсягів інформації, є не лише статистичні показники та рейтингові оцінки, а й слабкоструктуровані Big Data.

Інформаційна модель соціально-економічної системи, виходячи з положень загальної теорії систем [20, с. 23], має вигляд:

$$S_0 = S_1 \times S_2 \times \dots \times S_i \times \dots \times S_m, \quad (1)$$

де S_i — i -й ієрархічний рівень; m — кількість ієрархічних рівнів;

$$S_i = \langle M_i, P_i, R_i, X_i, Y_i, f_i, \varphi_i \rangle, \quad (2)$$

де M_i, P_i, R_i — множина реальних об'єктів, суб'єктів і підсистем i -го ієрархічного рівня; X_i, Y_i — множина внутрішніх і зовнішніх параметрів системи i -го ієрархічного рівня та зовнішнього середовища; φ_i, f_i — функціонали, що визначають взаємозв'язок відповідних параметрів на m рівнях у вигляді:

$$\varphi_i : X_i \rightarrow Y_i; \quad f_i : Y_i \rightarrow Y_{i-1}. \quad (3)$$

Таке представлення інформаційної моделі задачі дослідження складної системи передбачає її опис за допомогою інформації, яку сформовано різними способами та одержано з різних джерел.

За даною моделлю було опрацьовано інформаційну базу дослідження задачі прийняття рішення щодо вибору напрямів розвитку агропромислового комплексу України в умовах необхідності швидкої переорієнтації зовнішньоекономічної діяльності. Припинення експорту сільськогосподарської продукції у Російську Федерацію (частка російського ринку у структурі експорту продукції АПК скоротилася з 11 до 5 % [19]) призвело до необхідності перегляду структури експорту агропродовольчої продукції, пошуку нових ринків збуту і, відповідно, змін у структурі товарної продукції сільського господарства та харчової промисловості [19].

Виходячи з особливостей предметної області, описаних В.М. Ціхановською [19], задачу пошуку оптимальних співвідношень виробництва різних видів продукції АПК та її реалізації на світовому та внутрішньому ринках можна визначити як задачу дворівневого багатокритеріального вибору. У процесі формулювання задачі та під час її розв'язку використано методику багатокритеріального дворівневого аналізу, запропоновану П.С. Панкратьєвим та В.А. Шакіровим [12, с. 119], адаптовану до особливостей національної економіки та доповнену на етапах опрацювання альтернатив і цільових настанов аналізом неструктурованої текстової інформації.

Задача пошуку оптимальних співвідношень виробництва різних видів продукції АПК та її реалізації на світовому та внутрішньому ринках може бути сформульована наступним чином. Альтернативами першого рівня $A = \{a_1, a_2, \dots, a_n\}$ є потенційні імпортери української аграрної продукції. Множина альтернатив першого рівня оцінюється на множині критеріїв $C = C_1 \cup D$. Альтернативами другого рівня $B = \{b_{i1}, b_{i2}, \dots, b_{ik}\}$ є види продукції, що будуть реалізовані, обсяги реалізації та експортери конкретних видів сільськогосподарської продукції. Підмножина критеріїв $C_1 = \{c_1, c_2, \dots, c_p\}$ використовується для оцінки альтернатив виключно першого рівня, а множина критеріїв $D = \{d_1, d_2, \dots, d_s\}$ — як для першого, так і для другого рівнів. Кожному елементу множини A («покупцю») ставиться у відповідність множина альтернатив другого рівня (обсяги реалізації кожного виду сільськогосподарської продукції за окремим напрямком). Необхідно ранжувати альтернативи множини A , враховуючи багатокритеріальні оцінки альтернатив множини B . Складністю даної задачі є те, що на першому рівні для вибору альтернатив використовуються переважно якісні ознаки, а на другому — кількісні. Крім того, враховуючи, що в процесі відбору альтернатив можуть брати участь як вітчизняні, так і закордонні експерти, їхній набір, кількість і характеристики альтернатив можуть суттєво відрізнятися.

Методика використання засобів текстової аналітики, що запропонована у даній роботі, призначена для відбору та аналізу альтернатив, визначення цільової установки дослідження також передбачає послідовну реалізацію набору кроків, які охоплюють як попередню обробку текстової інформації, формування набору правил для аналізу, відбір найбільш значимих критеріїв і цілей, так і візуалізацію одержаних результатів.

Схему реалізації методики та використання засобів текстової аналітики можна представити у вигляді наступної послідовності кроків.

1. Сформулювати цілі та критерії відбору альтернатив, визначивши при цьому рівні аналізу та згрупувати альтернативи за рівнями аналізу.

2. Сформувати множину альтернатив першого рівня.
3. Сформувати множину альтернатив другого рівня.
4. Сформулювати множину оцінок альтернатив другого рівня за найбільш значимими критеріями.

5. Обрати метод багатокритеріального аналізу. Оскільки поточна ситуація в економіці України характеризується наявністю невизначеностей різного характеру, застосовується багатокритеріальна функція корисності альтернатив [12, с. 122]:

$$F(y) = \sum_{i=1}^n k_i f_i(y_i). \quad (4)$$

6. Побудувати однокритеріальні функції для оцінки корисності альтернатив, що відповідають критеріям для оцінки альтернатив першого та другого рівнів. При побудові функцій слід враховувати ризики для кожної оцінки корисності.

7. Виконати оцінку коефіцієнтів, що шкалюють альтернативи.

8. Виконати оцінку альтернатив другого рівня за допомогою функції корисності.

9. Сформувати множину з найкращих альтернатив другого рівня, відповідно до переваг особи, що приймає рішення.

10. Оцінити альтернативи першого рівня за множиною критеріїв, яка сформована виключно для першого рівня. За критеріями корисності та цінності для оцінки альтернатив першого і другого рівнів, на основі множини альтернатив, що сформована на кроці 9, відібрати набори альтернатив другого рівня.

11. Виконати попарні порівняння критеріїв із множини, яка сформована на кроці 10.

12. Виконати порівняння альтернатив першого рівня за критеріями, що представлені в множині оцінки критеріїв для першого та другого рівнів, і виконати перетворення оцінок корисності в оцінки відносної важливості.

13. Виконати парні порівняння альтернатив першого рівня за критеріями множини оцінки альтернатив виключно першого рівня.

14. Виконати багатокритеріальну оцінку альтернатив першого рівня.

15. Вибрати найкращу альтернативу першого рівня на основі аналізу відносної стійкості розв'язків наборів кращих альтернатив, що відібрані на кроці 10.

На етапах 1–4 запропоновано забезпечити автоматизоване опрацювання значних обсягів переважно текстової інформації за допомогою таких потужних сучасних інструментів як сімейство програмних засобів SAS Textual Analytics. Дані програмні засоби мають широкий спектр можливостей для аналізу структурованих і неструктурованих даних: це і SAS Web Crawler, SAS Search and Indexing, SAS Enterprise Content Categorization, SAS Ontology Management, SAS Text Miner та SAS Sentiment Analysis Studio.

У ході дослідження (етапи 1–4) дані для відбору альтернатив першого та другого рівнів пропонується видобувати з веб-простору, використовуючи SAS Web Crawler — програмне забезпечення, що сканує сторінки мережі Інтернет та використовує вбудовану структуру посилань, за допомогою якої веб-сторінки пов'язані одна з одною. Веб-краулеру необхідно вказати посилання на сторінку, з

якої необхідно розпочати сканування. Цю веб-сторінку ще називають точкою входу. У даному дослідженні точкою входу переважно є посилання безпосередньо на розділ новин певного веб-сайту. Звичайний веб-краулер починає свою роботу з дослідження точки входу і сканує розміщені на ній посилання, котрі, у свою чергу, посилаються на інші веб-сторінки. Віднайдені таким чином сторінки також скануються на наявність посилань на інші сторінки. Глибину такого пошуку можна обмежити шляхом встановлення значень певних параметрів. У ході парсингу веб-сторінок веб-краулер вилучає з них необхідний контент і зберігає його у вигляді файлів-документів.

Для визначення емоційного забарвлення текстової інформації, запропоновано побудувати множини правил для відбору позитивних і негативних уподобань для кожної частини мови.

Для проведення дослідження, яке передбачене на кроці 4, запропоновано використовувати інструменти SAS Sentiment Analysis Studio, що дозволяють аналізувати тональність тексту на основі побудованих правил, будувати моделі, в основу яких покладені правила та гібридні моделі (на основі статистичних моделей і моделей на основі правил).

Дослідження виконано на матеріалах щодо виробництва та реалізації основної експортної культури АПК України — кукурудзи. Перевірку працездатності та адекватності розробленої методики виконано у ході аналізу інтернет-джерел текстів щодо тематики виробництва в Україні окремих сільськогосподарських культур і пріоритетних напрямків їхньої реалізації на внутрішньому та світовому ринках для формування альтернатив першого та другого рівня задачі дворівневого багатокритеріального аналізу альтернатив у системі підтримки прийняття рішень розвитку агропромислового виробництва України.

Для аналізу корпусу текстів було використано SAS Textual Analytics Suit, що складається з SAS Content Categorization та SAS Sentimental Analysis Studio за допомогою яких розроблено таксономію із 776-ти правил. Одночасно з автоматизованим дослідженням виробництва та ринку кукурудзи, авторами виконувалось дослідження і без використання засобів текстової аналітики. Порівнюючи трудозатрати на обробку даних за двома способами, слід відзначити, що використання інструментів тестової аналітики компанії SAS значно знизило трудомісткість одержання експертного висновку: було витрачено 40 людино-годин проти 900 людино-годин, що були би потрібні без використання вказаних інструментів. Значними є обсяги опрацьованої інформації (табл. 1).

На рисунку представлено приклад побудови таксономії у SAS Content Categorization Studio, де відображено, як правило (на правій панелі) дозволяє на дистанції у 5 слів виконувати пошук словосполучень типу «постачання кукурудзи», «продаж кукурудзи».

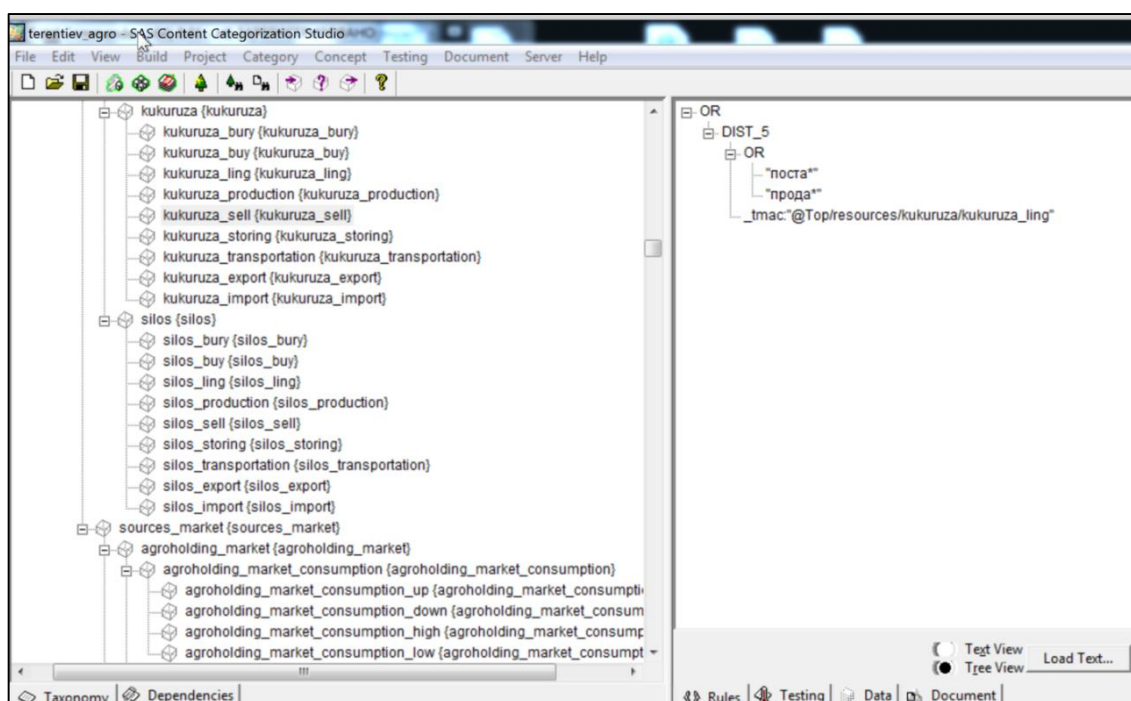
Для зручності побудови звітів отриманий набір даних, який по суті являє собою багатомірну OLAP-базу даних, було розкладено за чотирма рівнями.

Синтез правил щодо аналізу досліджуваної системи представлено на прикладі домена «сільське господарство» через виявлення станів «високий, низький, зростаючий або спадаючий рівень потенційно негативного або позитивного показника» з додатковим вилученням фактуальної словникової інформації (словник-

категоризатор країн, ієрархія понять з галузі сільського господарства з класифікатору КВЕД — табл. 2).

Таблиця 1. Перелік інтернет-джерел, що опрацьовані в ході дослідження

Назва інтернет-ресурсу	Кількість завантажених статей
http://latifundist.com/blog	5649
http://agroua.net/news/	1324
http://obozrevatel.com/blogs	2741
http://blog.liga.net	11008
http://blogs.lb.ua	3538
http://agroportal.ua/views/blogs	1038
http://www.agro24.com.ua/blog	116
http://blogs.pravda.com.ua/	23180



Результати побудови правил у програмі SAS Content Categorization Studio

Предметний домен S_0 «сільське господарство» має наступні складові S_{11} та S_{12} : <‘культури’, ‘ринки’> згідно моделі опису складної ієрархічної системи. Рівні S_{1i} складаються, у свою чергу, з інших рівнів S_{2j} , наприклад: S_{11} = <‘кукурудза’, ‘пшениця’, ‘ячмінь’, ‘зерно’, ‘гречка’, ‘соняшник’, ‘ріпак’ >.

Кожне джерело рівня S_{11} має над собою функціональні залежності, що перетворюють його кількісно чи якісно: $\phi_k(S_{1j}) \in$ <‘покупка’, ‘продаж’, ‘вирощування’,

‘селекція’, ‘транспортування’, ‘зберігання’, ‘переробка’, ‘використання’> та впливають на зовнішні параметри системи S_0 .

Таблиця 2. Назви груп КВЕД для генерації правил категоризатору

Тип	Н1	Н2	Н3	Н4	Текст	Ключові слова
ЛІ	А. Сільське господарство, лісове господарство та рибне господарство	01. Сільське господарство, мисливство та надання пов'язаних із ними послуг	01.1. Вирощування однорічних і дворічних культур	01.11. Вирощування зернових культур (крім рису), бобових культур і насіння олійних культур	Вирощування зернових культур, таких як:	Вирощування, вирощування зернових культур, зернова культура, культура
ЛІ	А. Сільське господарство, лісове господарство та рибне господарство	01. Сільське господарство, мисливство та надання пов'язаних із ними послуг	01.1. Вирощування однорічних і дворічних культур	01.11. Вирощування зернових культур (крім рису), бобових культур і насіння олійних культур	пшениця	пшениця
ЛІ	А. Сільське господарство, лісове господарство та рибне господарство	01. Сільське господарство, мисливство та надання пов'язаних із ними послуг	01.1. Вирощування однорічних і дворічних культур	01.11. Вирощування зернових культур (крім рису), бобових культур і насіння олійних культур	кукурудза (на зерно)	зерно кукурудза
ЛІ	А. Сільське господарство, лісове господарство та рибне господарство	01. Сільське господарство, мисливство та надання пов'язаних із ними послуг	01.1. Вирощування однорічних і дворічних культур	01.11. Вирощування зернових культур (крім рису), бобових культур і насіння олійних культур	сорго	сорго
ЛІ	А. Сільське господарство, лісове господарство та рибне господарство	01. Сільське господарство, мисливство та надання пов'язаних із ними послуг	01.1. Вирощування однорічних і дворічних культур	01.11. Вирощування зернових культур (крім рису), бобових культур і насіння олійних культур	ячмінь	ячмінь

Наступним кроком є синтез схеми, в якій за відповідними шаблонами зв'язуються показники з економічного словника (kpi) та рівні ієрархії предметного домену. Для визначення трендів до застосованих вище показників додані типові показники з економічного тезаурусу:

$$kpi_trends = kpi \times lvl \cap kpi \times dir, \quad (5)$$

де $kpi = \langle arg_j \rangle$, $arg_j \in \langle \text{споживання, вартість, дефіцит, попит, знижка, надлишок, інвестиції, вихід, ціна, квота, ризик, частка, ринкова вартість, субсидія, поставка, тариф, ставка податку, обсяг} \rangle$ та інші, $lvl = \langle arg_k \rangle$, $arg_k \in \langle \text{великий, низький} \rangle$, $dir = \langle arg_l \rangle$, $arg_l \in \langle \text{зріст, спад} \rangle$.

У результаті проведеного дослідження одержано наступні результати. Виявлено, що найбільш обговорюваними темами щодо тенденцій розвитку світового ринку аграрної продукції є наявність генно-модифікованих організмів (7,42 %), виробництво біодизеля (34,68 %), проблеми вирощування кукурудзи (55,87 %).

Щодо відбору альтернатив другого рівня, пов'язаних з учасниками ринку кукурудзи, то загалом на блогах, форумах і сайтах новин, що присвячені сільському господарству, найчастіше згадують про Україну — 21 %, Росію — 17 %, Європейський Союз — 17 %, США — 10 %, Польщу — 8,5 % та інші країни — 26,5 %. Найбільш важливими є альтернативи нарощування обсягів виробництва кукурудзи — 56 %, виробництво біопалива — 34 %, генетично модифіковані організми — 7 % та інші — 4 %.

У ході аналізу альтернатив значення цін і розвитку ринку сільськогосподарських культур, було виявлено, що розподіл пріоритетів змінювався протягом 2014–2016 маркетингових років. У 2015 р. структура альтернатив була наступною: ринкові розрахунки — 30 %; інвестиційна діяльність — 15 %; ринковий попит — 10 %; ринкова пропозиція — 10 %; розподіл ринку — 10 %; ефективність вирощування культур — 5 %. У 2016 р. структура дещо відрізняється — експертів більше цікавили питання про ринкові розрахунки (25 %) та інвестиційну діяльність (20 %).

Щодо емоційного забарвлення досліджуваної інформації, то слід відзначити зростання негативної тенденції щодо «квоти на продаж продукції» та «генно-модифіковані організми», позитивної — «кукурудза» та «Китай».

При порівнянні альтернатив, враховуючи наявність ризиків, використано метод аналізу ієрархій [8]. Для переведення оцінок корисності в оцінки порівнюваної важливості використовувався коефіцієнт переведення [12]:

$$K = \frac{c_{\max} - 1}{u_{\max} - u_{\min}}, \quad (6)$$

де u_{\max} та u_{\min} — відповідно, максимальна та мінімальна корисність альтернатив за певним критерієм; c_{\max} — максимальна оцінка відносної важливості критеріїв.

Максимальна оцінка важливості альтернатив визначається відповідно до шкали оцінок особою, що приймає рішення [12, с. 124].

Основними перевагами застосування текстової аналітики та дослідження емоційного забарвлення в процесі відбору альтернатив є можливість опрацювання значних обсягів неструктурованої інформації, що накопичена в різних джерелах, виявивши при цьому якнайбільшу кількість альтернатив за мінімальних трудовитрат особи, що приймає рішення, значно знизивши при цьому суб'єктивізм експертних оцінок.

При прийнятті рішення перевага віддається альтернативі з максимальним значенням математичного очікування та мінімальним стандартним відхиленням ваг. Ці значення розраховуються для кожного сценарію окремо.

Маючи кількісні характеристики альтернатив сценаріїв, на наступному рівні аналізу особа, що приймає рішення, має змогу візуалізувати профіль альтернативи сценарію у певний момент часу часового проміжку, порівняти між собою часові

перерізи альтернатив одного або декількох сценаріїв у визначений момент часу. Кожний фактор, що опрацьований за запропонованою методикою, описується певним інтегральним показником, до складу якого входять згруповані та впорядковані за важливістю для даного фактору лінгвістичні змінні у вигляді вербальних описів можливих варіантів розвитку подій. Це дає змогу формалізувати нечіткі уявлення щодо розвитку досліджуваного соціально-економічного процесу в рамках певного сценарію, який побудований на основі тенденцій, що виявлені експертами.

Отже, використання SAS Textual Analytics дозволило знизити вплив людського фактору на вибір альтернатив, значно пришвидшити опрацювання значних обсягів неструктурованої інформації з інтернет-джерел і сформулювати найбільш повне уявлення щодо розвитку ситуації, наявності ризиків та ознак кризових явищ, створити передумови для дослідження багатфакторних причинно-наслідкових залежностей, що можуть виникати серед різномірної та неструктурованої інформації, джерелом якої є відомості, одержувані з офіційних і неофіційних джерел.

Висновки

Дане дослідження виконано в рамках проекту НАТО «Безпека заради миру» NUKR.SFPP G4877 «Моделювання та попередження соціальних лих спричинених катастрофами та тероризмом», що виконується на базі Науково-навчального комплексу Інституту прикладного системного аналізу НТУУ «КПІ».

Враховуючи особливості розвитку соціально-політичної та фінансово-економічної ситуації в Україні, в дослідженні відзначено необхідність оптимізації системи розробки макроекономічних прогнозів і планів шляхом запровадження сценарного підходу, невід'ємною складовою якого є попередній аналіз предметної області, відбір альтернатив і формулювання цільових установок можливих варіантів розвитку подій. Використання засобів текстової аналітики як інструменту оптимізації підтримки прийняття рішень у задачах багатокритеріального вибору, розв'язуваних у рамках побудови сценаріїв розвитку соціально-економічних систем, як це було запропоновано у роботі, дозволяє найбільш повно відібрати альтернативи та визначити цільову настанову розвитку системи за допомогою запропонованої інформаційної моделі, яка узагальнює знання щодо структурних елементів системи, зв'язків між ними та впливу зовнішнього середовища.

Застосування запропонованої методики дозволить підвищити якість експертних висновків, у тому числі щодо забезпечення продовольчої безпеки країни та укріплення її позицій на світовому аграрному ринку. Крім того, її використання у роботі різних аналітичних центрів та органів державного управління дозволить не лише узгодити експертні оцінки, формалізувати нечіткі уявлення щодо розвитку національної економіки, її секторів і галузей, а й визначити орієнтири її розвитку на певних часових проміжках у межах розроблених сценаріїв.

1. *Офіційний сайт «Україна-НАТО»*. Програма НАТО «Наука заради миру та безпеки» [Електронний ресурс]. — Режим доступу: <http://ukraine-nato.mfa.gov.ua/ua/about-nato/science-for-peace>

2. *Дегтяр А.О.* Державно-управлінські рішення: інформаційно-аналітичне та організаційне забезпечення: монографія. / А.О. Дегтяр; Нац. акад. держ. упр. при Президентіві України, Харк. регіон. ін-т. — Харків: Вид-во ХАРІДУ НАДУ «Магістр», 2004. — 224 с.
3. *Дяченко Н.* Інформаційно-аналітичне забезпечення діяльності органів державної влади та місцевого самоврядування [Електронний ресурс] / Н. Дяченко // Публічне адміністрування: теорія та практика. — 2013. — Вип. 2. — Режим доступу: http://nbuv.gov.ua/j-pdf/Patp_2013_2_14.pdf. — Назва з екрану.
4. *Алексеев В.А.* Архітектура інтегрованої міжвідомчої інформаційної системи як композиція відомчих інформаційних систем / В.А. Алексеев, В.С. Терещенко // Проблеми програмування. — 2004. — № 2–3. — С. 397–408.
5. *Швиденко М.З.* Сучасні інформаційні технології моніторингу і аналізу стану інфраструктури аграрного ринку України / М.З. Швиденко // Науковий вісник Національного університету біоресурсів і природокористування України. — 2013. — Вип. 181(4). — С. 350–359.
6. *Стратегічні напрями розвитку сільського господарства України на період до 2020 року;* за ред. Ю.О. Лупенка, В.Я. Месель-Веселяка. — К.: ННЦ «ІАЕ», 2012. — 182 с.
7. *Харченко В.В.* До питання інформаційного забезпечення управління аграрного підприємництва / В.В. Харченко, В.Р. Онищук // Інноваційна економіка. — 2013. — № 5 (43). — С. 332–336.
8. *Саати Т.* Принятие решений. Метод анализа иерархий / Т. Саати; пер. с англ. Р.Г. Вачнадзе. — М.: Радио и связь, 1993. — 278 с.
9. *Згуровський М.З.* Технологічне передбачення економіки України на середньостроковому (до 2020 р.) і довгостроковому (до 2030 р.) часових горизонтах (за матеріалами наукової доповіді на засіданні Президії НАН України 4 листопада 2015 року) [Електронний ресурс] / М.З. Згуровський // Вісник Національної академії наук України. — 2016. — № 1. — С. 57–68. — Режим доступу: http://nbuv.gov.ua/UJRN/vnanu_2016_1_10
10. *Панкратова Н.Д.* Гибридный метод многокритериального оценивания альтернатив принятия решений / Н.Д. Панкратова, Н.И. Недашкова // Кибернетика и системный анализ. — 2014. — Т. 50, № 5. — С. 58–70.
11. *Ларичев О.И.* Теория и методы принятия решений / О.И. Ларичев. — М.: Логос, Университетская книга, 2006. — 296 с.
12. *Шакиров В.А.* Многокритериальный двухуровневый подход к выбору лучшей альтернативы в рамках слабоструктурированной проблемы / В.А. Шакиров, П.С. Панкратьев // Вестник ВГУ. Серия: Системный анализ и информационные технологии. — 2013. — № 2. — С. 118–127.
13. *Авдеева З.К.* Эвристический метод концептуальной структуризации знаний при формализации слабоструктурированных ситуаций на основе когнитивных карт / З.К. Авдеева, С.В. Коврига // Управление большими системами. — 2010. — № 31. — С. 6–34.
14. *Анализ данных и процессов: учеб. пособ.* / [А.А. Барсегян, М.С. Куприянов, И.И. Холод и др.]. — [3-е изд., перераб. и доп.] — СПб.: БХВ-Петербург, 2009. — 512 с.: ил. + CD-ROM — (Учебная литература для вузов). — ISBN 978-5-9775-0368-6.
15. *Survey of Text Mining I: Clustering, Classification, and Retrieval* / Ed. by M. W. Berry. — 2004. — Springer, 2003. — 261 p. — ISBN 0387955631.
16. *Aggarwal C.C. Mining Text Data* / C.C. Aggarwal, C. Zhai — Springer, 2012. — 527 p. — ISBN 9781461432234.
17. *Do Prado H.A. Emerging Technologies of Text Mining: Techniques and Applications* / Ed. by H.A. Do Prado, E. Ferneda. — Idea Group Reference, 2007. — 358 p. — ISBN 1599043734.

18. Панкратова Н.Д. Морфологічний аналіз: теорія, проблеми, застосування: навч. посіб. / Н.Д. Панкратова, І.О. Савченко. — К.: Наук. думка, 2015. — 245 с.
19. Ціхановська В.М. Оцінка стану та основні напрямки розвитку вітчизняного агропродовольчого ринку в умовах глобалізації [Електронний ресурс] / В.М. Ціхановська // Ефективна економіка. — 2016. — № 3. — Режим доступу: <http://www.economy.nauka.com.ua/?op=1&z=4819>
20. Панкратова Н.Д. Моделирование альтернатив сценариев процесса технологического предвидения / Н.Д. Панкратова, В.В. Севастьянов // Системні дослідження та інформаційні технології. — 2009. — N 1. — С. 22–35 .
21. *Text analytics using SAS Text Miner: course notes.* — NC.: SAS Institute, 2014. — 218 p.
22. Терентьев А.Н. SAS BASE: Основы программирования / А.Н. Терентьев, В.Н. Домрачев, Р.И. Костецкий. — К.: Эдельвейс, 2014. — 303 с. — ISBN 978-966-2748-49-9.

Надійшла до редакції 12.09.2016