

МЕТОД ПОБУДОВИ МОДЕЛІ ВЗАЄМОЗВ'ЯЗКУ «ЗМІ – СОЦІУМ – ДЕРЖАВНА ПОЛІТИКА» В ІНТЕРЕСАХ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ДЕРЖАВИ

Євген Добровольський

Національний технічний університет України «Київський політехнічний інститут», Україна



ДОБРОВОЛЬСЬКИЙ Євген Леонідович

Рік та місце народження: 1985 рік, м. Київ, Україна.

Освіта: Національний технічний університет України «Київський політехнічний інститут», 2009 рік.

Посада: здобувач кафедри інформаційної безпеки.

Наукові інтереси: інформаційна безпека.

Публікації: більше 10 наукових публікацій, серед яких монографії, наукові статті та тези наукових доповідей.

E-mail: evgdobr@gmail.com

Анотація. Для побудови систем оцінки стану соціальних явищ громадської активності, що спрямовані на забезпечення інформаційної безпеки держави, потребується розробка моделей, котрі дозволяють кількісно оцінювати взаємозв'язок «ЗМІ – Соціум – Державна Політика». Сучасні методи оцінки в переважній більшості базуються на ручній роботі експертів відповідної галузі та є мало автоматизовані. Це накладає обмеження на можливості існуючих систем підтримки прийняття рішень, наприклад, щодо можливості використання широкого кола вхідних даних та їх швидкодії. В роботі розроблено метод побудови моделі оцінки взаємозв'язку «ЗМІ – Соціум – Державна Політика», що дозволяє автоматизувати розрахунок параметрів її математичного рівняння та базується на результатах застосування контент-методів аналізу інформаційних потоків ЗМІ. Модель дозволяє автоматизувати процес оцінювання та розрахунок кількісного прогнозу змін настроїв громади, а також мінімізувати вплив людського чинника на отримувані результати. Використання даної моделі дозволяє отримувати оцінки в якісній та кількісній формах.

Ключові слова: інформаційна безпека держави, модель громадської активності, ЗМІ, соціум, державна політика, контент-аналіз.

Вступ

Інформаційні технології, що базуються на класичних методах контент-аналізу, використовуються в сьогоденні для розв'язання широкого кола складних проблем у сфері національної безпеки. Зокрема великої популярності набувають методи та моделі на основі дослідження інформаційних потоків засобів масової інформації (ЗМІ) [2,5,6]. Однак, досі розрахунок якісних і кількісних вимірів соціальних явищ виконується політологами в більшості випадків «в ручному режимі», тобто лише експерти, виконуючи аналіз останніх новин політичного та суспільного життя, можуть надавати висновки щодо зміни настрою суспільства, визначати взаємозв'язки різноманітних подій[10]. Недоліком даного методу є те, що на отримані результати дослідження накладають свій відбиток політичні вподобання дослідника[9,10]. Саме тому є актуальною задача з розробки новітнього інструменту (методу побудови моделі) аналізу соціальних явищ, що базується на використанні сучасних методик контент-дослідження інформаційних потоків [9].

Аналіз існуючих досліджень

Аналіз існуючих методологій вивчення та моніторингу соціальних явищ показав, що їх

головними недоліками є відсутність можливості щодо автоматизації та значний вплив людського чинника, наявні методи не дозволяють отримувати результати у форматі «он-лайн» [9,10]. Зроблено висновок щодо необхідності побудови інструменту, що дозволить усунути дані недоліки.

Дослідження джерел інформаційних потоків ЗМІ Україномовного простору Інтернет показало, що вони містять вичерпний опис соціальних явищ[6]. Зазначені джерела володіють наступними властивостями[6,10]:

- актуальність інформації;
- достатній рівень об'єктивності, що досягається великою кількістю інформаційних джерел;
- висока швидкість розповсюдження інформації;
- подання інформації у цифровому вигляді, що придатний до обробки комп'ютерними системами.

Зроблено висновок щодо можливості використання інформаційних джерел українського простору всесвітньої мережі Інтернет для дослідження соціальних явищ.

На основі дослідження сучасних теоретичних моделей взаємозв'язку «ЗМІ – Соціум – Державна

політика»[10] зроблено висновки щодо можливих шляхів їх практичної реалізації та автоматизації з можливістю використання результатів задля забезпечення інформаційної безпеки держави.

Метою роботи є розробка методу побудови моделі взаємозв'язку «ЗМІ - Соціум - Державна політика» в інтересах забезпечення інформаційної безпеки держави.

Основна частина дослідження

Отже однією з головних потреб експертів інформаційної безпеки держави в сьогоденні є отримання інструментарію, що дозволяє розраховувати оцінки взаємозв'язку «ЗМІ - Соціум - Державна політика» в автоматизованому режимі[2]. Для побудови такого рішення пропонується

використати системи контент-аналізу, що дозволяють визначати кількісні та якісні характеристики інформаційних потоків ЗМІ Інтернет. В багатьох роботах відзначено переваги використання таких інформаційних джерел.

Розробку методу побудови моделі залежності «ЗМІ - Соціум - Державна політика»[10] виконаємо на прикладі статистичних даних передвиборної кампанії кандидата в Президенти в 2009+2010 роках табл.1. Реакція соціуму на дії влади в даному випадку описана як рейтинг політичної особистості[3], а для опису подій довкола політичної особисті використано кількості статей з позитивним, негативним та нейтральним відгуком[7,8].

Вхідні статистичні данні

Таблиця 1

Дата	Рейтинг політичної особистості, % Y	Кількість статей з позитивним відгуком про політичну особистість X ₁	Кількість статей з негативним відгуком про політичну особистість X ₂	Кількість статей з нейтральним відгуком X ₃
21.12.2009	30,2	38	52	21
22.12.2009	30,3	37	50	22
23.12.2009	30,5	39	55	20
24.12.2009	30,4	37	53	23
25.12.2009	31,3	42	52	24
26.12.2009	31,5	43	54	24
27.12.2009	31,7	43	51	23
28.12.2009	32	44	53	26
29.12.2009	32,2	48	50	27
30.12.2009	32,5	53	49	25
31.12.2009	33,3	54	51	28
08.01.2010	33,3	53	51	29
09.01.2010	33,5	55	53	27
10.01.2010	33,7	55	57	31
11.01.2010	33,4	52	49	32
12.01.2010	33,9	57	48	33
13.01.2010	34,2	61	51	32
14.01.2010	34,2	63	47	35
15.01.2010	34,3	63	48	36
16.01.2010	34,4	64	47	37

Дата	Рейтинг політичної особистості, % Y	Кількість статей з позитивним відгуком про політичну особистість X ₁	Кількість статей з негативним відгуком про політичну особистість X ₂	Кількість статей з нейтральним відгуком X ₃
17.01.2010	34,5	65	42	35
18.01.2010	34,4	63	43	39
19.01.2010	34,6	71	40	40
20.01.2010	34,6	70	41	41
21.01.2010	34,7	72	39	43
22.01.2010	34,8	74	38	42
23.01.2010	34,9	75	37	45
24.01.2010	34,7	72	40	47
25.01.2010	35	76	37	46
26.01.2010	35,2	77	37	49
27.01.2010	35,3	79	35	51
28.01.2010	35,4	85	36	47
29.01.2010	35,6	89	32	54
30.01.2010	35,7	92	32	55
31.01.2010	35,9	95	33	57
01.02.2010	36	112	30	59
02.02.2010	36,2	117	31	61
03.02.2010	36,3	119	29	58
04.02.2010	36,2	115	27	65
05.02.2010	36,4	117	28	67

Метод побудови моделі передбачає виконання наступних кроків:

Крок 1. Виконання контент-дослідження. За допомогою автоматизованої системи контент-аналізу

виконано відбір унікальних статей, що описують події довкола певної політично особистості[7]. Пошуковий запит, що був використаний для отримання інформації, містив лише прізвище

політичної особи (зادля аполітизації статті прізвище не зазначено). За допомогою цієї ж автоматизованої системи виконано кластаризацію статей за їх тональним забарвленням на: позитивні, негативні та нейтральні[8].

Крок 2. Розрахунок оцінки рівняння регресії моделі. Рівняння множинної регресії моделі, що буде описувати явище взаємозв'язку «ЗМІ-Соціум-Державна політика» на основі визначених вище даних (табл. 1), можливо представити у вигляді:

$$Y = f(\beta, X) + \varepsilon,$$

де $X = (X_1, X_2, \dots, X_m)$ – вектор незалежних змінних, відповідно до табл.1: $m = \overline{1,3}$; β – вектор параметрів математичного рівняння моделі; ε – відхилення теоретичної моделі (помилка моделі); Y – залежна змінна, що описується моделлю (рейтинг політичної особистості). Оскільки $m = \overline{1,3}$, то теоретичне лінійне рівняння множинної регресії може бути переписано в більш простому вигляді:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon,$$

де β_0 – вільний параметр, що визначає Y , у випадку, коли $\forall i = \overline{1,3}: X_i = 0$. Емпіричне рівняння множинної регресії представляється у вигляді:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \varepsilon,$$

де b_0, b_1, b_2, b_3 – оцінки теоретичних значень коефіцієнтів $\beta_0, \beta_1, \beta_2, \beta_3$ регресії, а ε – оцінка відхилення ε . Надалі в рамках даного кроку методу виконується розрахунок оцінок коефіцієнтів регресії, що відповідно до методу найменших квадратів (МНК) визначається за формулою[1]:

$$s = (X^T X)^{-1} X^T Y.$$

Для даних з табл. 1 вектор оцінок коефіцієнтів

$$\text{регресії становить: } s = (X^T X)^{-1} X^T Y = \begin{pmatrix} 24,15 \\ 0,0168 \\ 0,0766 \\ 0,14 \end{pmatrix}.$$

Відповідно оцінка рівняння регресії приймає вигляд: $Y = 24,15 + 0,0168 \cdot X_1 + 0,0766 \cdot X_2 + 0,14 \cdot X_3$.

Крок 3. Розрахунок матриці парних коефіцієнтів. Надалі в рамках другого кроку дослідження пропонується виконати розрахунок матриці парних коефіцієнтів кореляції[1], відтак кількість спостережень дорівнює кількості днів дослідження і становить $n = 42$ (табл.1), кількість незалежних змінних для моделі становить 3, а кількість регресорів дорівнює кількості невідомих коефіцієнтів з урахуванням одиничного вектора (що слугує для введення коефіцієнта β_0). Тоді матриця, незалежних змінних X має розмірність 42×5 , а парні коефіцієнти та параметри кореляції можуть бути розраховані наступним чином (табл. 2, 3 та 4):

Парні коефіцієнти кореляції

Таблиця 2

Змінні	$\sum_{i=0}^n x_i$	$\bar{x} = \frac{\sum_{i=0}^n x_i}{n}$	$\sum_{i=0}^n y_i$	$\bar{y} = \frac{\sum_{i=0}^n y_i}{n}$	$\sum_{i=0}^n x_i y_i$	$\overline{xy} = \frac{\sum_{i=0}^n x_i y_i}{n}$
y і x ₁	2976	70,86	1430,4	34,06	103124,9	2455,35
y і x ₂	1777	42,31	1430,4	34,06	59893,8	1426,04
y і x ₃	1693	40,31	1430,4	34,06	58675,5	1397,04
x ₁ і x ₂	1777	42,31	2976	70,86	116381	2770,98
x ₁ і x ₃	1693	40,31	2976	70,86	135004	3214,38
x ₂ і x ₃	1693	40,31	1777	42,31	66107	1573,98

Параметри кореляції

Таблиця 3

Змінні	Дисперсія $D(x) = \frac{\sum_{i=0}^n x_i^2}{n} - \bar{x}^2$	Дисперсія $D(y) = \frac{\sum_{i=0}^n y_i^2}{n} - \bar{y}^2$	Середньо квадратичне відхилення $s(x) = \sqrt{D(x)}$	Середньо квадратичне відхилення $s(y) = \sqrt{D(y)}$	Коефіцієнт кореляції $r_{yx} = \frac{\bar{x} \cdot \bar{y} - \overline{xy}}{s(x) \cdot s(y)}$
y і x ₁	645,88	3,35	25,41	1,83	0,91
y і x ₂	88,69	3,35	9,42	1,83	-0,86
y і x ₃	207,88	3,35	14,42	1,83	0,92
x ₁ і x ₂	88,69	645,88	9,42	25,41	-0,95
x ₁ і x ₃	207,88	645,88	14,42	25,41	0,98
x ₂ і x ₃	207,88	88,69	14,42	9,42	-0,97

Таблиця 4

Матриця парних коефіцієнтів кореляції

	y	x ₁	x ₂	x ₃
y	1	0,91	-0,86	0,92
x ₁	0,91	1	-0,95	0,98
x ₂	-0,86	-0,95	1	-0,97
x ₃	0,92	0,98	-0,97	1

Аналіз інформації першого рядка табл. 4 дозволяє виконати відбір тих параметрів моделі, котрі мають бути включені в модель множинної кореляційної залежності, а ті параметри, для яких $|r_{y, x_i}| < 0,5$ надалі виключають з моделі як такі, що не

є вагомими. Як бачимо в рамках нашого дослідження та побудови моделі такі параметри відсутні.

Крок 4. Розрахунок t- статистик Стьюдента.

Даний крок методу передбачає розрахунок значень t-статистик: коефіцієнти регресії, для яких значення t-статистики за модулем більше від критичного значення, вважаються значущими[1]. Так для r_{y, x_1} t-статистика розраховується за формулою: $t_{\text{спост}} = r_{y, x_i} \frac{\sqrt{n-m-1}}{\sqrt{1-r_{y, x_i}^2}}$, де $m = 1$ – кількість факторів у рівнянні

регресії, тоді $t_{\text{спост}} = 0,91 \frac{\sqrt{42-1-1}}{\sqrt{1-0,91^2}} = 13,63$. За таблицею

Стьюдента знаходимо $t_{\text{крит}} \left(n - m - 1; \frac{\alpha}{2} \right) = t_{\text{крит}}(40; 0,025) = 2,021$. Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то гіпотезу про рівність нулю коефіцієнта кореляції відхилено, коефіцієнт кореляції є статистично-значимий. Аналогічно розрахуємо спостережувані значення t -статистики для $r_{y x_2}$: $t_{\text{спост}} = 0,86 \frac{\sqrt{42-1-1}}{\sqrt{1-0,86^2}} = 10,9$. Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то коефіцієнт кореляції є статистично-значимий. Значення t -статистики для $r_{y x_3}$: $t_{\text{спост}} = 0,92 \frac{\sqrt{42-1-1}}{\sqrt{1-0,92^2}} = 14,64$. Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то коефіцієнт кореляції є також статистично-значимий.

Таким чином можливо зробити висновок, що зв'язки між y і x_1 , y і x_2 , y і x_3 є значимими, при чому найбільший вплив на результативну ознаку Y виявляє фактор x_3 для якого $r_{y x_3} = 0,92$. Дійсно, численні дослідження показали, що людина сприймає краще «суху» інформацію, у випадках передбачуваного вибору, а розміщені у тексті позитивно-негативні акценти розсіюють увагу й зміщають пріоритети.

Крок 5. Розрахунок приватних коефіцієнтів кореляції. Більш точні характеристики мір зв'язку допомагають визначити приватні коефіцієнти кореляції. Дані коефіцієнти відрізняються від звичайних коефіцієнтів лінійної парної кореляції тим, що вони вимірюють парну кореляцію відповідних ознак (y і x_i) за умови, що вплив на них інших факторів (x_j) фіксований. На основі таких коефіцієнтів стає можливим робити висновок щодо включення змінних в регресійну модель: якщо значення коефіцієнту є незначним, то зв'язок між цим фактором і результативною змінною дуже слабкий, тому фактор можна виключити з результативної моделі. Відтак розрахуємо значимість приватних коефіцієнтів кореляції за формулою[1]:

$$r_{y x_i / x_j} = \frac{r_{y x_i} - r_{y x_j} r_{x_i x_j}}{\sqrt{(1 - r_{y x_j}^2)(1 - r_{x_i x_j}^2)}}$$

Звідси $r_{y x_1 / x_2} = \frac{0,91 - 0,92 \cdot 0,98}{\sqrt{(1 - 0,92^2)(1 - 0,98^2)}} = 0,12$ (тіснота зв'язку низька), а спостережувані значення t -статистики розрахуємо за формулою $t_{\text{спост}} = r_{y x_1 / x_2} \frac{\sqrt{n-k-1}}{\sqrt{1 - r_{y x_1 / x_2}^2}}$, де $k = 1$ - число «зафіксованих»

факторів моделі. Звідси $t_{\text{спост}} = 0,55 \frac{\sqrt{42-1-2}}{\sqrt{1-0,55^2}} = 4,06$. За таблицю Стьюдента $t_{\text{крит}} \left(n - m - 1; \frac{\alpha}{2} \right) = t_{\text{крит}}(39; 0,025) = 2,021$. Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то гіпотеза про рівність нулю коефіцієнта кореляції відхиляється, коефіцієнт кореляції є статистично-значимий. Зв'язок y і x_1 за умови, що x_2 увійде в модель, знизився.

Для $r_{y x_1 / x_3}$: $t_{\text{спост}} = 0,12 \frac{\sqrt{42-1-2}}{\sqrt{1-0,12^2}} = 0,73$. Оскільки $t_{\text{спост}} < t_{\text{крит}}$, то гіпотеза про рівність нулю коефіцієнта кореляції вірна, тобто коефіцієнт кореляції статистично не значимий, а $r_{y x_1 / x_3} = \frac{-0,86 - 0,91 \cdot (-0,95)}{\sqrt{(1 - 0,91^2)(1 - 0,95^2)}} = -0,036$ - тіснота зв'язку низька. Зв'язок y і x_1 за умови, що x_3 увійде в модель, знизився.

Для $r_{y x_2 / x_1}$: $t_{\text{спост}} = 0,036 \frac{\sqrt{42-1-2}}{\sqrt{1-0,036^2}} = 0,22$.

Оскільки $t_{\text{спост}} < t_{\text{крит}}$, то гіпотеза про рівність нулю коефіцієнта кореляції вірна, тобто коефіцієнт кореляції статистично не значимий. Далі $r_{y x_2 / x_3} = \frac{-0,86 - 0,92 \cdot (-0,97)}{\sqrt{(1 - 0,92^2)(1 - 0,97^2)}} = 0,24$ - тіснота зв'язку низька, зв'язок y і x_2 за умови, що x_1 увійде в модель, знизився.

Значимість коефіцієнта кореляції $r_{y x_2 / x_3} = \frac{0,92 - 0,91 \cdot 0,98}{\sqrt{(1 - 0,91^2)(1 - 0,98^2)}} = 0,35$ - тіснота зв'язку не сильна. $t_{\text{спост}} = 0,24 \frac{\sqrt{42-1-2}}{\sqrt{1-0,24^2}} = 1,57$, а оскільки $t_{\text{спост}} < t_{\text{крит}}$, то ухвалюємо гіпотезу про рівність нулю коефіцієнта кореляції тобто, коефіцієнт кореляції статистично - не значимий. Зв'язок y і x_2 за умови, що x_3 увійде в модель, знизився.

Для $r_{y x_3 / x_1} = 0,35$ тіснота зв'язку помірна, $t_{\text{спост}} = 0,35 \frac{\sqrt{42-1-2}}{\sqrt{1-0,35^2}} = 2,36$. Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то гіпотеза про рівність нулю коефіцієнта кореляції відхиляється, коефіцієнт кореляції є статистично-значимий. Зв'язок y і x_3 за умови, що x_1 увійде в модель, знизився.

Значимість коефіцієнта кореляції $r_{y x_3 / x_2} = 0,35$, а значення t -статистики $t_{\text{спост}} = 0,64 \frac{\sqrt{42-1-2}}{\sqrt{1-0,64^2}} = 5,23$. Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то гіпотезу про рівність нулю коефіцієнта кореляції відхилено, коефіцієнт кореляції статистично - значимий. Зв'язок y і x_3 за умови, що x_2 увійде в модель, знизився.

За результатом даного кроку можна зробити висновок, що при побудові регресійного рівняння моделі «ЗМІ-Соціум-Державна політика» слід застосовувати всі три фактори x_1, x_2, x_3 .

Крок 6. Розрахунок моделі регресії в стандартному масштабі. Модель регресії в стандартному масштабі припускає, що всі значення досліджуваних ознак переводяться в стандарти (стандартизовані значення). Система нормальних рівнянь у такому випадку має вигляд:

$$\begin{cases} r_{y x_1} = \beta_1 + r_{x_1 x_2} \beta_2 + r_{x_1 x_3} \beta_3 \\ r_{y x_2} = r_{x_2 x_1} \beta_1 + \beta_2 + r_{x_2 x_3} \beta_3 \\ r_{y x_3} = r_{x_3 x_1} \beta_1 + r_{x_3 x_2} \beta_2 + \beta_3 \end{cases}$$

Для дослідження моделі «ЗМІ-Соціум-Державна політика» та даних з табл.4 така система прийме вигляд:

$$\begin{cases} 0,907 = \beta_1 - 0,948 \beta_2 + 0,977 \beta_3 \\ -0,865 = -0,948 \beta_1 + \beta_2 - 0,968 \beta_3 \\ 0,918 = 0,977 \beta_1 - 0,968 \beta_2 + \beta_3 \end{cases}$$

Розв'язком даної системи рівнянь є:

$$\begin{cases} \beta_1 = 0,233 \\ \beta_2 = 0,394 \\ \beta_3 = 1,072 \end{cases}$$

І тому стандартизована форма рівняння регресії приймає вигляд:

$$y = 0,233 x_1 + 0,394 x_2 + 1,072 x_3$$

Крок 7. Статистичний аналіз рівняння регресії. Даний крок виконання методу містить статистичний аналіз розрахованого рівняння регресії, а саме перевірку значимості рівняння та його коефіцієнтів, дослідження абсолютних та відносних помилок апроксимації. Підготовка інформація, що потребується для виконання такого

аналізу наведена в табл.5. Безпосередньо крок даного методу передбачає розрахунок наступних величин, що свідчать про достовірність побудованої моделі[1]:

$$1. \text{ Середня помилка апроксимації: } A = \frac{\sum |\varepsilon|}{n}.$$

$$100\% = 1,767\%.$$

$$2. \text{ Значення оцінки дисперсії: } s_e^2 = (Y - X*Y(X))^T(Y - X*Y(X)) = 20.44.$$

$$3. \text{ Незміщена оцінка дисперсії: } s^2 = \frac{1}{n-m-1}.$$

$$s_e^2 = \frac{1}{n-m-1} \cdot 20,44 = 0,54.$$

$$4. \text{ Оцінка середньоквадратичного відхилення (стандартна помилка для оцінки } Y): S = \sqrt{s^2} = 0,73.$$

5. Оцінка ковариаційної матриці вектора:

$$k = S(X^T X)^{-1} = \begin{pmatrix} 14,76 & -0,00377 & -0,22 & -0,13 \\ -0,00377 & 0,000607 & 4,3 \cdot 10^{-5} & -0,00102 \\ -0,22 & 4,3 \cdot 10^{-5} & 0,00317 & 0,00193 \\ -0,13 & -0,00102 & 0,00193 & 0,00306 \end{pmatrix}.$$

Діагональні елементи даної матриці визначають дисперсії параметрів моделі.

6. Приватні коефіцієнти еластичності, що показують на скільки відсотків в середньому змінюється Y зі зміною X_i на один відсоток від свого середнього значення за умови фіксованих інших факторів моделі. Такі приватні коефіцієнти еластичності розраховуються за формулою: $E = b_i \frac{\bar{x}_i}{\bar{y}}$. Для досліджуваної моделі «ЗМІ-Соціум-Державна політика»:

$|E_1| = 0,0349 < 1$ - його вплив на результативну ознаку Y незначний.

$|E_2| = 0,0951 < 1$ - його вплив на результативну ознаку Y незначний.

$|E_3| = 1,0100 > 1$ - його вплив на результативну ознаку Y значний.

7. Долю впливу кожного X_i в загальному результативному значенні Y визначають за коефіцієнтами роздільної детермінації: $d_i^2 = r_{yx_i} \cdot \beta_i$. Для досліджуваної моделі «ЗМІ-Соціум-Державна політика» вони становлять:

$$d_1^2 = 0,21,$$

$$d_2^2 = -0,34,$$

$$d_3^2 = 0,98.$$

При цьому виконується рівність: $\sum_{i=1}^3 d_i^2 = R^2 = 0,85$. Отже зв'язок між ознакою Y та факторами X для моделі «ЗМІ-Соціум-Державна політика» сильний.

8. Коефіцієнт детермінації R^2 . При значенні R^2 близькому до 1 рівняння регресії добре описує фактичні дані, а при значенні R^2 , що близьке до 0, вважається, що рівняння регресії не досить добре описує фактичні дані[4]. В рамках даного дослідження $R^2 = 0,85$.

9. Перевірка значущості параметрів множинного рівняння регресії на основі t-статистики за формулами:

$$t_{\text{табл}}(n - m - 1; \frac{\alpha}{2}) = t_{\text{табл}}(38; 0,025) = 2,021,$$

$$t_i = \frac{b_i}{S_{b_i}}.$$

Стандартна помилка коефіцієнта регресії b_0 : $S_{b_0} = \sqrt{14,76} = 3,84$, $t_0 = \frac{24,15}{3,84} = 6,28 \geq 2,021$. Статистична значимість коефіцієнта регресії b_0 підтверджується.

Стандартна помилка коефіцієнта регресії b_1 : $S_{b_1} = \sqrt{0,000607} = 0,0246$, $t_1 = \frac{0,0168}{0,0246} = 0,68 \leq 2,021$. Статистична значимість коефіцієнта регресії b_1 не підтверджується.

Стандартна помилка коефіцієнта регресії b_2 : $S_{b_2} = \sqrt{0,000317} = 0,0563$, $t_2 = \frac{0,0766}{0,0563} = 1,36 \leq 2,021$. Статистична значимість коефіцієнта регресії b_2 не підтверджується.

Стандартна помилка коефіцієнта регресії b_3 : $S_{b_3} = \sqrt{0,000306} = 0,0553$, $t_2 = \frac{0,14}{0,0553} = 2,46 \leq 2,021$. Статистична значимість коефіцієнта регресії b_3 підтверджується.

Розраховано довірчі інтервали коефіцієнтів регресії для рівня надійності 95%:

$$(b_i - t_i S_{b_i}; b_i + t_i S_{b_i}):$$

$$b_0: (16,38; 13,91),$$

$$b_1: (-0,033; 0,0666),$$

$$b_2: (-0,0373; 0,19),$$

$$b_3: (0,0241; 0,25).$$

Таблиця 5

Підготовча інформація

Y	Y(x) (значення теоретичної моделі)	$\varepsilon = Y - Y(x)$ (незміщена помилка)	ε^2	$(Y - Y_{\text{ср}})^2$	$ \frac{\varepsilon}{Y} $
30,2	31,62	-1,42	2,02	14,88	0,0471
30,3	31,59	-1,29	1,66	14,12	0,0425
30,5	31,73	-1,23	1,52	12,65	0,0404
30,4	31,95	-1,55	2,41	13,37	0,0511
31,3	32,1	-0,8	0,64	7,6	0,0255
31,5	32,27	-0,77	0,59	6,54	0,0244
31,7	31,9	-0,2	0,0405	5,56	0,00635
32	32,48	-0,48	0,23	4,23	0,015
32,2	32,45	-0,25	0,0638	3,45	0,00785
32,5	32,19	0,31	0,0974	2,42	0,0096
33,3	32,77	0,53	0,29	0,57	0,016
33,3	32,89	0,41	0,17	0,57	0,0125
33,5	32,8	0,7	0,49	0,31	0,0209
33,7	33,65	0,0498	0,00248	0,13	0,00148

Закінчення таблиці 5

33,4	33,12	0,28	0,0767	0,43	0,00829
33,9	33,27	0,63	0,4	0,0247	0,0187
34,2	33,43	0,77	0,6	0,0204	0,0226
34,2	33,56	0,64	0,41	0,0204	0,0186
34,3	33,78	0,52	0,28	0,059	0,0153
34,4	33,85	0,55	0,3	0,12	0,016
34,5	33,21	1,29	1,66	0,2	0,0373
34,4	33,8	0,6	0,36	0,12	0,0174
34,6	33,84	0,76	0,58	0,29	0,0219
34,6	34,04	0,56	0,32	0,29	0,0163
34,7	34,19	0,51	0,26	0,41	0,0147
34,8	34,01	0,79	0,62	0,55	0,0227
34,9	34,36	0,54	0,29	0,71	0,0155
34,7	34,81	-0,11	0,0119	0,41	0,00315
35	34,51	0,49	0,24	0,89	0,014
35,2	34,94	0,26	0,07	1,31	0,00752
35,3	35,09	0,21	0,0451	1,54	0,00601
35,4	34,72	0,68	0,46	1,8	0,0192
35,6	35,43	0,17	0,0276	2,38	0,00467
35,7	35,62	0,0799	0,00639	2,7	0,00224
35,9	36,02	-0,12	0,0142	3,4	0,00331
36	36,35	-0,35	0,12	3,77	0,00962
36,2	36,78	-0,58	0,34	4,59	0,016
36,3	36,25	0,0486	0,00236	5,03	0,00134
36,2	36,98	-0,78	0,61	4,59	0,0216
36,4	37,37	-0,97	0,93	5,49	0,0265
36,6	37,17	-0,57	0,32	6,47	0,0156
36,6	37,53	-0,93	0,87	6,47	0,0255
		0	20,44	140,5	0,74

Крок 8. Загальна перевірка якості рівняння множинної регресії. Оцінка значимості рівняння множинної регресії виконується за допомогою перевірки гіпотези щодо рівності нулю коефіцієнта детермінації [4], який розрахований за даними генеральної сукупності: R^2 або $b_1 = b_2 = \dots = b_m = 0$. Для такої перевірки гіпотези використовують F -критерій Фішера, при цьому обчислюють фактичне значення F -критерію, через коефіцієнт детермінації R^2 . За таблицями розподілу Фішера-Снедекора знаходять критичне значення F -критерію ($F_{кр}$), для цього задають рівнем значимості α (для нашого дослідження ми використаємо $\alpha = 0,05$) та двома значеннями ступенів волі $k_1 = m$ та $k_2 = n - m - 1$. Тоді:

$$R^2 = 1 - \frac{s_e^2}{\sum(y-\bar{y})^2} = 1 - \frac{20,44}{140,5} = 0,87.$$

Більш об'єктивною оцінкою є скоректований коефіцієнт детермінації:

$$\bar{R}^2 = 1 - (1 - R^2) \cdot \frac{n - 1}{n - m - 1} = 1 - (1 - 0,87^2) \cdot \frac{41 - 1}{42 - 3 - 1} = 0,85.$$

Перевірка гіпотези про загальну значимість – гіпотези про одночасну рівність нулю всіх коефіцієнтів регресії при пояснюючих змінних: $H_0: \beta_1 = \beta_2 = \beta_3 = 0$. Перевірка цієї гіпотези здійснюється за допомогою F -статистики розподілу Фішера, якщо $F < F_{кр} = F_{\alpha; n-m-1}$, то немає підстав для відхилення гіпотези H_0 :

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m} = 74,45.$$

Табличне значення $F_{кр}$ при визначених ступенях волі становить: $F_{кр}(3;38)=2,84$. Оскільки фактичне значення $F > F_{кр}$, то коефіцієнт детермінації статистично значимий й рівняння регресії статистично надійно.

Висновки

Отже запропонований в даній роботі метод побудови моделі взаємозв'язку «ЗМІ-Соціум-Державна політика» базується на виконанні наступних кроків дослідження:

Крок 1. Виконання контент-дослідження.

Крок 2. Розрахунок оцінки рівняння регресії моделі.

Крок 3. Розрахунок матриці парних коефіцієнтів.

Крок 4. Розрахунок t -статистик Стьюдента.

Крок 5. Розрахунок приватних коефіцієнтів кореляції.

Крок 6. Статистичний аналіз рівняння регресії

Крок 7. Розрахунок моделі регресії в стандартному масштабі.

Крок 8. Загальна перевірка якості рівняння множинної регресії.

Завдяки розробленому методу та побудованій моделі стає можливим автоматизований процес моделювання та прогнозування оцінок взаємозв'язку між владою та громадою, що реалізується засобами масової інформації. Вихідні дані даної моделі можуть бути використані в системах підтримки

прийняття рішень та системах побудови правил прийняття рішень в інтересах забезпечення інформаційної та громадської безпеки держави.

Література

[1] Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998. – 650 с.

[2] Горбулін В.П. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія / В.П. Горбулін, О.Г. Додонов, Д.В. Ланде. - К.: Інтертехнологія, 2009. - 164 с.

[3] Звіти Київський міжнародний інститут соціології [Електронний ресурс] - Режим доступу <http://www.kiis.com.ua/?lang=ukr&cat=reports>

[4] Згуровський М.З. Основи системного аналізу [Текст] : підручник / М.З. Згуровський, Н.Д. Панкратова ; за ред. М.З. Згуровського. - К. : ВНУ, 2007. - 543 с. : табл. - (Інформатика). - Бібліогр.: С. 517-531.

[5] Качинський А.Б. Засади системного аналізу безпеки складних систем / Качинський А.Б. - К.: ДП «НВЦ «Євроатлантик-форм», 2006. - 336 с. (Формування і реалізація державної політики

управління процесами європейської та євроатлантичної інтеграції України).

[6] Ланде Д.В. Поиск знаний в Internet. Профессиональная работа / Ланде Д.В. - М.: Издательский дом «Вильямс», 2005 - 182 с.

[7] Науково-практичний журнал Національного університету оборони України «Сучасні Інформаційні технології у сфері безпеки та оборони» №2(8) 2010, Добровольський Є., Качинський А. «Застосування сучасних інформаційних технологій для дослідження політичної активності громади», стор. 81-93.

[8] Науково-практичний журнал Національної академії служби безпеки України «Інформаційна безпека» №1 (5) 2011, Добровольський Є., Качинський А., Ланде Д., «Вивчення статистичних характеристик громадської активності населення, як індикатора загроз національній безпеці України», стор. 59-67.

[9] Парсонс В. Публічна політика (вступ до теорії й практики аналізу) / Вейн Парсонс, - К.: «Видавничий дім «Києво-Могилянська академія»», 2006.

[10] Штромайер Г. Політика і мас-медіа (переклад з нім.) / Штромайер Г. - К.: Вид. дім «Києво-могилянська академія», 2008 - 303 с.

УДК 321.01:351 (045)

Добровольський Е.Л. Метод построения модели взаимосвязи «СМИ – Социум – Государственная Политика» в интересах обеспечения информационной безопасности государства

Аннотация. Для построения систем оценки состояния социальных явлений общественной активности, направленные на обеспечение информационной безопасности государства, требуется разработка методов и моделей, которые позволяют количественно оценивать взаимосвязь «СМИ – Социум – Государственная Политика». Современные методы оценки в подавляющем большинстве основаны на ручной работе экспертов в соответствующей области и являются мало автоматизированы. Это накладывает ограничения на возможности существующих систем, например, относительно возможности использования широкого круга входных данных и их быстродействия. В работе разработан метод построения модели оценки взаимосвязи «СМИ – Социум – Государственная Политика», что позволяет автоматизировать расчет параметров ее математического уравнения, и базируется на результатах применения контент-методов анализа информационных потоков СМИ. Модель позволяет автоматизировать процесс оценивания и расчет количественного прогноза изменений настроений общества, а также минимизировать влияние человеческого фактора на получаемые результаты. Использование данной модели позволяет получать оценки в качественной форме.

Ключевые слова: информационная безопасность государства, модель общественной активности, СМИ, социум, государственная политика, контент-анализ.

Dobrovolskyi Ye. Method for relationship model «MASS MEDIA – SOCIETY – STATE POLICY» constructing for information security of the state ensuring

Abstract. For building systems condition assessment of social phenomena social activity aimed at ensuring information security of the state, requires the development of methods and models that allow us to quantitatively evaluate the relationship between «Mass media – Society – Public Policy». Modern methods of assessment and the vast majority are based on manual work of experts in the relevant field and are not automated. This imposes constraints on existing systems, for example, about the possibility of using a wide range of input data and their performance. In this paper, a method of constructing a model of the relationship «Mass media – Society – Public Policy» that allows you to automate the calculation of the parameters of its mathematical equations and is based on the results of the application of content analysis methods of information flow media. The model allows to automate the process of estimation and calculation of the quantitative forecast of changes in attitudes of society, and to minimize the influence of human factor on the results. The use of this model allows obtaining estimates in a qualitative way.

Key words: information security of the state, model of social activity, media, society, public policy, content analysis.

Отримано 1 червня 2015 року, затверджено редколегією 24 червня 2015 року