Список літератури

1. *Редіко Л. Раз* крок, два крок. // Схемотехніка. – 2001. – №6 – №11.
2. Довідник по теорії автоматичного управління. / під ред. *А.А. Красовського*. – М.: Наука, 1997.
3. *Копилов І.П.* Математичне моделювання електричних машин. – М.: Висш. шк., 2002.
4. *Солоха А.А.* Математическая модель шагового двигателя // – Кострома.– 2004. – Т.5.
5. *Нейдорф р.А.* Завдання квазіоптимальної швидкодії управління шаговим двигуном. // Математичні методи в техніці і технологіях: матеріали XVII междунар. научн. конф. – Казань. – 2005. – Т.2.
6. *Нейман Л.Р. Демірчан К.С.* Теоретичні основи електротехніки. – Л.: Енергія, 1987.
7. *Пий Ан.* Сполучення ПК із зовнішніми пристроями. – М.: ДМК Прес; Спб.: Пітер, 2004.
8. *Растрыгин л.А.* Системы экстремального управления. – М.: Наука, 1984.

Досліджується можливість реалізації оптимального або квазіоптимального по швидкодії управління кроком ротора двигуна за рахунок спеціальної організації багатоканального генератора імпульсів. У статті розглядаються два варіанти управління шаговим двигуном, в одному з них оцінка ефективності проводиться по математичній моделі шагового двигуна, в другому за експериментальними даними.

Исследуется возможность реализации оптимального или квазиоптимального по быстродействию управления шагом ротора двигателя за счет специальной организации многоканального генератора импульсов. В статье рассматриваются два варианта управления шаговым двигателем, в одном из них оценка эффективности проводится по математической модели шагового двигателя, во втором по экспериментальным данным.

Marketability by the step of rotor of engine optimum or квазиоптимального on a fast-acting management is probed due to the special organization of multichannel pulser. Two variants of foot-pace engine management are examined in the article, in one of them the estimation of efficiency is conducted on the mathematical model of foot-pace engine, in the second from experimental data.

Рецензент: Єрохін В.Ф.
Надіслано 04.11.2010

УДК 21.973-018.2.я7

Кузнецов Г.В., Сушко С.О. (НГУ)

МАТРИЦЯ ПЕРЕХІДНИХ ЙМОВІРНОСТЕЙ ДЛЯ МОДЕЛІ УКРАЇНСЬКОГО ТЕКСТУ, ГЕНЕРОВАНОГО СТАЦІОНАРНИМ ДЖЕРЕЛОМ МАРКОВСЬКІ ЗАЛЕЖНИХ БУКВ

Імовірнісні моделі джерел відкритих повідомлень використовуються у криптографії безпосередньо в алгоритмах дешифрування, а також для розрізнення відкритого осмисленого тексту і випадкових послідовностей. Вибір підхідної моделі для опису джерела відкритого тексту, зазвичай, здійснюється криптоаналітиком залежно від властивостей конкретного шифру. Відкритий текст – це джерело випадкових послідовностей, а текст, породжений цим джерелом, – імовірнісний аналог мови.

Очевидно, що розробка подібних моделей потребує дослідження інформаційно-статистичних властивостей мовних об'єктів. Для багатьох мов такі роботи проводились вже давно, зокрема, англійську досліджував автор і класик теорії інформації К.Шеннон [1]. Радянський академік Піотровський Р.Г. із співробітниками одержали багато цікавих інформаційно-статистичних параметрів російської та інших мов колишнього СРСР [2 – 4]. Деякі дослідження статистичних властивостей української мови, проведені в Інституті мовознавства ім. О.О.Потебені НАН України, на жаль, були неповними через використання недостатньо широкого обсягу текстового матеріалу, часто зводились до порівняння різних стилів деяких письменників.

У даній роботі за модель відкритого тексту обрано однозв'язний ланцюг Маркова і скінченною кількістю станів, який визначається матрицею П перехідних ймовірностей вектором π початкового розподілу ймовірностей:

$$\text{П}=(p(a/b)); \quad 0 \le a < m-1;\ 0 \le b < m-1;$$
$$v = (v(0), v(1), ..., v(m-1)),$$

де $v(b)$ – ймовірність появи букви b на першій позиції випадкового тексту.

Ймовірність появи випадкового тексту $(a_0, a_1, ..., a_{n-1})$ дорівнює

$$P(a_0, a_1, ..., a_{n-1}) = v(a_0) \cdot p(a_1/a_0) \cdot p(a_2/a_1) \cdot ... \cdot p(a_{n-1}/a_{n-2}).$$

Перехідні ймовірності та початковий розподіл знаків у випадковому тексті задовольняють умови:

$$p(a/b) \ge 0,\ v(b) \ge 0 \text{ для всіх } a, b \in Z_m;$$
$$v(0) + v(1) + ... + v(m-1) = 1;$$
$$p(0/b) + p(1/b) + ... + p((m-1)/b) = 1 \text{ для всіх } b \in Z_m.$$

Вектор $w = (w(0), w(1), ..., w(m-1))$ стаціонарного розподілу знаків у випадковому тексті можна знайти з системи рівнянь:

$$w(a) = \sum_{t=0}^{m-1} w(b) \cdot p(a/b), \quad a \in Z_m.$$

Марковський процес є прикладом стохастичного процесу з внутрішньо символьною залежністю (пам'яттю). Значення букви a_i у відкритому тексті $(a_0, a_1, ..., a_{n-1})$ залежіть від значення попередньої букви a_{i-1}. Залежність сусідніх станів визначається перехідними ймовірностями $p(a_i/a_{i-1})$ – елементами матриці П.

Базуючись на власних статистичних дослідженнях українських літературних текстів, автори поставили за мету сформувати матрицю П перехідних ймовірностей для української мови. Для цього було використано вибірку з текстів з двільних сторінок україномовних сайтів. Тексти належали до п'яти стилів сучасної української мови: розмовно-побутового, художнього, наукового, публіцистичного та ділового. Загальний об'єм текстових даних склав 900123 слів (приблизно 625Мб). Вихідні тексти перед розрахунками піддали спеціальній обробці (вилучено імена, прізвища, власні назви, літери інших мов, апостроф, дати, наявні в тексті фрагменти окремих слів, цифрові дані, скорочення одиниць вимірювання фізичних величин, абревіатуру, переноси слів, розділові знаки, відступи на початках абзаців). Усі слова в текстах було розділено одним пробілом, ототожнено великі й малі букви. Підрахунок кількості повторюваності в тексті букв, біграм проводився за спеціально розробленою програмою, що базується на використанні бінарного дерева пошуку.

У результаті статистичної обробки матеріалу обчислено ймовірності появи кожного символу української абетки в тексті (табл. 1) та ймовірності переходу $p(a/b)$ (матриця П, табл.2). Розподіл рівноваги $w = (w(0), w(1), ..., w(m-1))$ для марковського ланцюга наведено у табл. 3.

Таблиця 1. Середньостатистичні частоти букв та пропуску між словами в українській мові

␣	0,138	І	0,044	Д	0,027	Г	0,013	Ж	0,007
О	0,086	Р	0,043	Л	0,027	Ч	0,011	Ю	0,008
Н	0,068	Е	0,042	П	0,025	Х	0,011	Є	0,005
А	0,064	С	0,037	З	0,020	Ї	0,010	Щ	0,004
И	0,055	К	0,033	Я	0,019	Ц	0,010	Ф	0,003
В	0,046	М	0,029	Ь	0,016	Ш	0,005	Ґ	0,000
Т	0,045	У	0,027	Б	0,013	Й	0,009		

Таблиця 2. Матриця перехідних ймовірностей

	А	Б	В	Г	Ґ	Д	Е	Є	Ж
А	0,000	0,015	0,047	0,021	0	0,030	0,000	0,023	0,013
Б	0,104	0,000	0,003	0,005	0	0,002	0,099	0,014	0,000
В	0,148	0,001	0,003	0,002	0	0,017	0,037	0,000	0,009
Г	0,154	0,000	0,002	0,000	0	0,000	0,014	0,000	0,000
Ґ	0,000	0,000	0,000	0,000	0	0,000	0,000	0,000	0,000
Д	0,080	0,011	0,019	0,004	0	0,005	0,071	0,000	0,029
Е	0,008	0,010	0,022	0,015	0	0,038	0,000	0,000	0,025
Є	0,000	0,000	0,024	0,000	0	0,046	0,000	0,000	0,000
Ж	0,115	0,002	0,001	0,000	0	0,008	0,309	0,000	0,002
З	0,256	0,045	0,067	0,012	0	0,018	0,022	0,001	0,000
И	0,000	0,005	0,055	0,006	0	0,013	0,000	0,004	0,002
І	0,007	0,011	0,103	0,008	0	0,107	0,000	0,009	0,008
Ї	0,000	0,000	0,025	0,000	0	0,000	0,000	0,000	0,000
Й	0,000	0,020	0,006	0,001	0	0,017	0,000	0,000	0,003
К	0,103	0,000	0,033	0,000	0	0,000	0,010	0,000	0,000
Л	0,133	0,000	0,000	0,002	0	0,006	0,109	0,001	0,000
М	0,139	0,000	0,002	0,000	0	0,000	0,081	0,000	0,003
Н	0,174	0,000	0,001	0,001	0	0,006	0,065	0,001	0,000
О	0,000	0,052	0,105	0,070	0	0,059	0,003	0,003	0,020
П	0,047	0,000	0,001	0,000	0	0,005	0,128	0,000	0,000
Р	0,149	0,002	0,009	0,012	0	0,003	0,130	0,001	0,009
С	0,023	0,000	0,037	0,000	0	0,000	0,026	0,000	0,000
Т	0,144	0,000	0,050	0,000	0	0,000	0,126	0,002	0,001
У	0,007	0,008	0,078	0,021	0	0,058	0,000	0,023	0,014
Ф	0,135	0,001	0,000	0,000	0	0,000	0,103	0,000	0,000
Х	0,062	0,000	0,008	0,001	0	0,007	0,001	0,000	0,000
Ц	0,021	0,000	0,003	0,000	0	0,000	0,260	0,002	0,000

Ч	0,221	0,000	0,001	0,000	0	0,000	0,143	0,000	0,000
Ш	0,067	0,000	0,013	0,000	0	0,000	0,178	0,000	0,000
Щ	0,044	0,000	0,000	0,000	0	0,000	0,154	0,000	0,000
Ь	0,000	0,017	0,000	0,000	0	0,000	0,000	0,000	0,000
Ю	0,001	0,005	0,059	0,000	0	0,037	0,000	0,034	0,000
Я	0,000	0,000	0,032	0,025	0	0,026	0,000	0,009	0,009
␣	0,023	0,032	0,102	0,015	0	0,056	0,015	0,005	0,004

	З	**И**	**І**	**Ї**	**Й**	**К**	**Л**	**М**	**Н**
А	0,020	0,000	0,000	0,018	0,018	0,044	0,090	0,054	0,114
Б	0,000	0,056	0,083	0,000	0,000	0,004	0,079	0,012	0,068
В	0,004	0,165	0,115	0,000	0,000	0,010	0,029	0,002	0,066
Г	0,001	0,022	0,056	0,002	0,000	0,004	0,073	0,003	0,020
Ґ	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Д	0,003	0,081	0,071	0,001	0,002	0,021	0,033	0,006	0,109
Е	0,031	0,000	0,000	0,003	0,012	0,065	0,033	0,057	0,168
Є	0,002	0,000	0,000	0,036	0,000	0,025	0,000	0,108	0,021
Ж	0,002	0,104	0,030	0,000	0,000	0,008	0,063	0,001	0,134
З	0,000	0,023	0,024	0,000	0,000	0,018	0,022	0,037	0,110
И	0,015	0,000	0,000	0,004	0,045	0,051	0,029	0,061	0,061
І	0,039	0,000	0,000	0,051	0,050	0,023	0,049	0,011	0,074
Ї	0,003	0,000	0,000	0,073	0,005	0,017	0,001	0,016	0,093
Й	0,000	0,000	0,000	0,000	0,000	0,006	0,023	0,015	0,098
К	0,000	0,138	0,071	0,000	0,000	0,000	0,033	0,000	0,016
Л	0,005	0,174	0,111	0,000	0,000	0,007	0,001	0,000	0,002
М	0,000	0,160	0,087	0,001	0,000	0,015	0,007	0,000	0,016
Н	0,002	0,132	0,090	0,000	0,001	0,015	0,000	0,000	0,090
О	0,028	0,000	0,000	0,042	0,002	0,026	0,036	0,081	0,055
П	0,004	0,045	0,085	0,000	0,000	0,000	0,059	0,000	0,005
Р	0,004	0,132	0,075	0,003	0,001	0,029	0,001	0,040	0,033
С	0,000	0,058	0,029	0,000	0,000	0,038	0,052	0,002	0,051
Т	0,000	0,156	0,071	0,000	0,001	0,019	0,003	0,001	0,029
У	0,011	0,000	0,000	0,001	0,003	0,090	0,054	0,031	0,026
Ф	0,002	0,003	0,137	0,000	0,000	0,003	0,003	0,002	0,005
Х	0,000	0,018	0,064	0,000	0,000	0,001	0,002	0,001	0,041
Ц	0,000	0,031	0,416	0,014	0,009	0,000	0,010	0,002	0,007

Ч	0,000	0,209	0,028	0,000	0,000	0,009	0,004	0,000	0,288
Ш	0,000	0,208	0,078	0,000	0,000	0,048	0,035	0,008	0,081
Щ	0,000	0,032	0,016	0,000	0,000	0,000	0,000	0,000	0,000
Ь	0,000	0,000	0,000	0,000	0,001	0,269	0,000	0,010	0,135
Ю	0,001	0,000	0,000	0,000	0,000	0,001	0,000	0,001	0,001
Я	0,034	0,000	0,000	0,000	0,001	0,131	0,012	0,054	0,041
_	0,078	0,000	0,035	0,009	0,015	0,040	0,008	0,047	0,076

	О	**П**	**Р**	**С**	**Т**	**У**	**Ф**	**Х**	**Ц**
А	0,000	0,016	0,045	0,053	0,064	0,007	0,002	0,027	0,032
Б	0,107	0,000	0,044	0,034	0,010	0,224	0,000	0,017	0,001
В	0,113	0,008	0,005	0,035	0,006	0,022	0,000	0,001	0,003
Г	0,489	0,002	0,091	0,001	0,003	0,037	0,000	0,000	0,000
Ґ	0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000
Д	0,199	0,019	0,031	0,012	0,005	0,079	0,000	0,004	0,001
Е	0,018	0,016	0,209	0,033	0,037	0,001	0,004	0,015	0,015
Є	0,000	0,006	0,011	0,002	0,174	0,000	0,000	0,000	0,000
Ж	0,030	0,000	0,000	0,000	0,001	0,072	0,000	0,000	0,001
З	0,048	0,024	0,031	0,007	0,001	0,038	0,000	0,001	0,001
И	0,000	0,013	0,030	0,079	0,057	0,000	0,002	0,096	0,023
І	0,007	0,002	0,015	0,049	0,034	0,000	0,001	0,007	0,002
Ї	0,000	0,000	0,001	0,002	0,003	0,000	0,000	0,053	0,001
Й	0,059	0,003	0,002	0,090	0,014	0,001	0,000	0,000	0,001
К	0,287	0,000	0,074	0,013	0,049	0,069	0,000	0,000	0,024
Л	0,122	0,000	0,000	0,001	0,003	0,030	0,001	0,000	0,000
М	0,159	0,035	0,004	0,010	0,000	0,089	0,000	0,000	0,001
Н	0,170	0,000	0,000	0,027	0,030	0,030	0,012	0,000	0,008
О	0,002	0,019	0,075	0,067	0,025	0,001	0,002	0,008	0,009
П	0,277	0,000	0,279	0,008	0,002	0,016	0,000	0,000	0,003
Р	0,218	0,001	0,000	0,016	0,017	0,049	0,000	0,008	0,002
С	0,040	0,055	0,000	0,001	0,312	0,047	0,005	0,006	0,010
Т	0,114	0,000	0,063	0,004	0,011	0,049	0,000	0,000	0,002
У	0,000	0,027	0,040	0,043	0,060	0,000	0,000	0,006	0,002
Ф	0,423	0,000	0,088	0,001	0,001	0,087	0,000	0,000	0,000
Х	0,124	0,000	0,025	0,001	0,011	0,017	0,000	0,000	0,000
Ц	0,006	0,000	0,000	0,001	0,065	0,003	0,000	0,000	0,000

Ч	0,039	0,000	0,000	0,000	0,001	0,027	0,000	0,000	0,000
Ш	0,105	0,004	0,000	0,001	0,059	0,064	0,000	0,000	0,000
Щ	0,712	0,000	0,000	0,000	0,000	0,012	0,000	0,000	0,000
Ь	0,094	0,000	0,000	0,119	0,025	0,000	0,000	0,000	0,003
Ю	0,000	0,000	0,003	0,001	0,279	0,000	0,000	0,001	0,002
Я	0,000	0,000	0,009	0,009	0,047	0,000	0,000	0,020	0,001
␣	0,033	0,116	0,037	0,069	0,059	0,028	0,012	0,007	0,022

	Ч	**Ш**	**Щ**	**Ь**	**Ю**	**Я**	**␣**
А	0,016	0,007	0,001	0,000	0,014	0,001	0,207
Б	0,003	0,004	0,000	0,000	0,001	0,001	0,024
В	0,006	0,002	0,001	0,000	0,000	0,017	0,172
Г	0,003	0,000	0,000	0,000	0,001	0,001	0,020
Ґ	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Д	0,007	0,001	0,000	0,005	0,001	0,013	0,074
Е	0,010	0,004	0,001	0,000	0,004	0,003	0,143
Є	0,003	0,001	0,000	0,000	0,066	0,000	0,476
Ж	0,003	0,000	0,000	0,000	0,000	0,002	0,111
З	0,002	0,002	0,000	0,020	0,000	0,007	0,161
И	0,032	0,006	0,010	0,000	0,000	0,007	0,294
І	0,023	0,017	0,001	0,000	0,009	0,020	0,261
Ї	0,002	0,000	0,000	0,000	0,000	0,000	0,704
Й	0,000	0,016	0,000	0,000	0,000	0,000	0,626
К	0,000	0,001	0,004	0,000	0,000	0,000	0,073
Л	0,000	0,000	0,000	0,159	0,038	0,084	0,011
М	0,028	0,000	0,000	0,000	0,000	0,008	0,152
Н	0,002	0,008	0,000	0,019	0,007	0,082	0,026
О	0,016	0,006	0,001	0,000	0,025	0,002	0,161
П	0,000	0,001	0,000	0,000	0,026	0,003	0,006
Р	0,000	0,012	0,000	0,005	0,005	0,017	0,018
С	0,000	0,001	0,000	0,094	0,001	0,090	0,020
Т	0,000	0,000	0,000	0,111	0,005	0,011	0,026
У	0,023	0,004	0,002	0,000	0,021	0,002	0,346
Ф	0,003	0,000	0,000	0,000	0,000	0,000	0,003
Х	0,000	0,000	0,000	0,000	0,000	0,000	0,613
Ц	0,000	0,000	0,000	0,073	0,020	0,048	0,007

Ч	0,007	0,000	0,000	0,000	0,000	0,007	0,015
Ш	0,000	0,001	0,000	0,001	0,001	0,000	0,047
Щ	0,000	0,000	0,000	0,000	0,000	0,000	0,029
Ь	0,000	0,031	0,000	0,000	0,001	0,003	0,291
Ю	0,075	0,000	0,001	0,000	0,030	0,000	0,465
Я	0,006	0,000	0,001	0,000	0,007	0,000	0,526
␣	0,015	0,004	0,020	0,000	0,000	0,018	0,000

Таблиця 3. Розподіл рівноваги для марковського ланцюга

а	0,061	**є**	0,005	**к**	0,037	**с**	0,038	**ш**	0,005
б	0,014	**ж**	0,008	**л**	0,026	**т**	0,052	**щ**	0,004
в	0,047	**з**	0,021	**м**	0,031	**у**	0,024	**ь**	0,016
г	0,013	**и**	0,050	**н**	0,080	**ф**	0,005	**ю**	0,008
ґ	0,000	**і**	0,034	**о**	0,078	**х**	0,012	**я**	0,019
д	0,028	**ї**	0,011	**п**	0,031	**ц**	0,010	**–**	0,137
е	0,035	**й**	0,009	**р**	0,039	**ч**	0,012		

Очевидно, для кращої апроксимації української мови необхідно обчислити ймовірності переходів, що залежать більше, ніж від однієї попередньої букви абетки. Отримані результати можуть бути корисними в практичній діяльності криптоаналітиків та для модельних досліджень українських текстів у сфері криптології.

Список літератури

1. Шеннон К. Работы по теории информации и кибернетике. – М.: Наука, 1973. – 832 с.
2. Белоногов Г.Г., Фролов Г.Д. Эмпирические данные о распределении букз в русской письменной речи // В сборнике «Проблемы передачи кибернетики». – 1963. – Вып. 9. – С. 287 – 305.
3. Пиотровский Р.Г. Информационные измерения языка. – Л.: Наука, 1968. – 116 с.
4. Статистика речи. Сборник. Отв. редактор Пиотровский Р.Г. – Л.: Наука, 1968. – 299 с.
5. Статистичні та структурні лінгвістичні моделі. – К. Наукова думка, 1966. – 371 с.
6. Вирт Н.К. Алгоритмы+структуры данных = программы. – М.: Мир, 1985. – 410 с.

У роботі наведено результати досліджень інформаційно-статистичних властивостей української мови, що застосовуються в алгоритмах крипто аналізу.

Ключові слова: модель тексту, марковськи залежні букви, матриця перехідних ймовірностей.

В работе приведены результаты исследований информационно-статистических свойств украинского языка, которые используются в алгоритмах криптоанализа.

Ключевые слова: модель текста, марковски зависимые буквы, матрица переходных вероятностей.

The paper presents the results of research information and statistical properties of the Ukrainian language used in cryptanalysis algorithms.

Рецензент: Хорошко В.О.
Надійшла 27.05.2010