УДК 004.932.2

Калініченко Ю.В.
ДЗ «Луганський національний університет
імені Тараса Шевченка»

АДАПТИВНИЙ МЕТОД РОЗПІЗНАВАННЯ ДИНАМІЧНОГО АЛФАВІТНО-ЦИФРОВОГО РЯДУ

У статті проведено аналіз актуальності та методів розпізнавання динамічних алфавітно-цифрових рядів. Запропонована модифікація методу умовних випадкових полів з прихованими станами. Вона дозволяє підвищити дискримінативні можливості даної моделі. Проведено дослідження моделі на алгоритмічному рівні. Встановлено, що дана модифікація може бути алгоритмічно реалізована і має сублінійну тимчасову складність.

***Ключові слова**: динамічний алфавітно-цифровий ряд, розпізнавання образів, метод умовних випадкових полів із прихованими станами, генеративна імовірнісна модель, дискримінативна імовірнісна модель.*

Постановка проблеми. В останні роки в сучасному суспільстві у зв'язку з інтенсивним розвитком віртуальних технологій, а також з дослідженнями в галузі природничих людино-машинних інтерфейсів, зріс інтерес розробників і дослідників до обробки інформації динамічних алфавітно-цифрових рядів.

Процес розпізнавання динамічного алфавітно-цифрового ряду є частиною вирішення прикладних задач, таких як розпізнавання та оцифрування текстів вивісок, оголошень, розпізнавання автомобільних номерів, читання показань лічильників обліку комунальних послуг, визначення номерів будинків та назв вулиць, провулків, кварталів тощо для створення карт (Google Street View).

В якості генератора алфавітно-цифрового ряду може виступати той чи інший машинний інтерфейс, що змінює оптичну конфігурацію в просторі. Динамічний алфавітно-цифровий ряд – послідовність букв алфавіту, цифр або символів, що змінюють один одного у часі конфігурації.

Динаміка володіє більшою інформаційною ємністю, тому задача розпізнавання динамічних даних не вирішена в повному обсязі і є перспективною та актуальною науково-технічною задачею. Інтерфейси, що засновані на представлені динамічних даних, володіють більшою інтерактивністю та швидкістю отримання інформації від/до користувача. Тому виникає певна специфіка для користування даними сучасними інтерфейсами загального призначення на основі динамічних алфавітно-цифрових рядів для систем віртуальної реальності. Це визначає ряд вимог до методів розпізнавання даних та отримання інформації:

1) робота в режимі реального часу,

2) можливість створення спеціальних баз запису даних,

3) здатність реєстрації зміни конфігурації генератора алфавітно-цифрових рядів з заданою частотою,

4) збереження високої точності розпізнавання тощо.

Постановка завдання. Існуючі методи володіють тими чи іншими функціональними обмеженнями, у зв'язку із цим дана галузь потребує дослідження та розробки нових методів розпізнавання.

Метою даного дослідження є розробка адаптивного методу розпізнавання динамічних алфавітно-цифрових рядів, здатного відповідати поставленим вимогам.

У завдання дослідження входить аналіз існуючих методів, вибір найбільш відповідного прототипу, його модифікація на теоретичному рівні та алгоритмічне дослідження.

Виклад основного матеріалу дослідження. На сучасному етапі можна виділити генеративні і дискримінативні імовірнісні моделі, з точки зору точності розпізнавання та можливості роботи в режимі реального часу.

Генеративні моделі мають широке застосування як у розпізнаванні динамічних даних, так і в інших галузях, таких як автоматичне розпізнавання мови чи аналіз природних мов. Серед генеративних моделей розпізнавання виділяють метод прихованих марковських моделей, як найбільш популярний. Однак дані моделі володіють таким недоліком, що полягає в припущенні про те, що спостереження незалежні між собою. Під час розпізнавання динамічних алфавітно-цифрових даних це припущення не можна вважати коректним [1].

У дискримінативних моделях цей недолік відсутній. Дані моделі враховують контекст здійснення динаміки. Крім цього, вони володіють перевагою, що дозволяє забезпечити більш високу точність, ніж генеративні моделі, при меншому обсязі навчальної вибірки, що є актуальним для створення баз даних [1]. Серед дискримінативних моделей виділяють метод умовних випадкових полів та його модифікації. Це є найбільш перспективним методом розпізнавання динамічних даних.

Однак у роботі з великими базами даних має місце втратить точність розпізнавання [1]. Даний недолік можна усунути шляхом додавання нелінійної функції сигмоїдальної активації.

Алгоритмічне дослідження модифікованої моделі. Послідовність $X = \{x_1, \ldots , x_T\}$ довжини T впорядкована в часі та виступає в якості вхідних значень. Її довжина залежить від вихідних даних та може сильно варіюватися. Кожен елемент послідовності представляє собою характеристичний вектор $x_e \in M^D$ розмірності D, який описує одномоментну конфігурацію генерацію алфавітно-цифрового ряду.

Кожна послідовність алфавітно-цифрового ряду відображається в мітку класу $y \in Y$, де Y – множина визначених раніше значень цього ряду.

Для розпізнавання потрібно зіставити послідовність, за якою ведеться спостереження, з міткою, відповідного класу. Згідно з визначенням умовних випадкових полів із прихованими станами [1; 2] умовна ймовірність події визначається як:

$$p(y \,|\, X;W) = \frac{1}{Z(X;W)} \sum_h e^{F(y,h,X;W)}.$$

W – вектор параметрів моделі h H $h \in H$, де H являє собою множину прихованих станів, $Z(X;W)$ – коефіцієнт нормалізації, а $F(y, h, X; W)$ – функція ознак.

Коефіцієнт нормалізації розраховується наступним чином:

$$Z(x;W) = \sum_y \sum_h e^{F(y,h,x;w)}.$$

У свою чергу, функція ознак має вигляд:

$$F(y,h,X;W) = \sum_t f^1(h,x,t;w) + \sum_t f^2(y,h,t;w) + $$
$$+ \sum_t f^3(y,h,t,t+1;w)$$

Якщо I – індикаторна функція, така, що $I(x) = \begin{cases} 1, & якщо\ x - істина \\ 0, & якщо\ x - хибність \end{cases}$, тоді другий та третій вираз у функції ознак можемо записати таким чином:

$$f^2(h,x,t;w) = w_{y,h} I(y = y') I(h_t = h') \text{ та}$$

$$f^2(h,x,t;w) = w_{y,h} I(y = y') I(h_t = h'),$$

де $y' \in Y$, h', $h'' \in H$, а $w_{y,h}$, $w_{y,h,h}$ – відповідні ваги.

Для того, щоб виділити в просторі ознак складної форми, необхідно ввести безліч функцій активації G та функцію $\psi_g(x,t;W)$ для визначення f^1 як

$$\psi_g(x,t;W) = \frac{1}{|r(x_t)|} \sum_{x' \in r(x_t)} g\left(\sum_d w_{g,d} x'_d\right).$$

Сигмоїдальна функція активації $\mathrm{g(x)} = \frac{1}{(1 + \exp(-x))}$ виступає в якості функції активації. У свою чергу, отримуємо:

$$f^1(h,x,t;w) = \sum_{g \in G} w_{g,h} I(h_t = h') \psi_g(x,t;W).$$

Таким чином, функції активації створюють додатковий проміжок між прихованими змінних і даними, за якими ведеться спостереження, тим самим абстрагують необхідні дані.

Перевага даної модифікації помітна під час порівняння зі звичайним визначенням функції:

$$f^1(h,x,t;w) = \frac{1}{|r(x_t)|} \sum_{x'} \sum_{g \in G} w_{h,d} I(h_t = h') x'_d.$$

Виходячи з визначення, етап навчання відсутній та замість нього використовується лінійна комбінація ознак x'_d та ваг $w_{h,d}$. Використання нелінійності в моделі дозволяє підвищити дискримінаційні здібності і тим самим збільшити точність розпізнавання складних просторово-часових закономірностей, таких як динамічні алфавітно-цифрові ряди [3].

Представимо параметричний вектор у моделі, як $W = \{w_{g,h}, w_{g,d}, w_{y,h}, w_{y,h,h}\}$. Розмірність цього вектора складає $GH+GD+YH+YHH$, де G – кількість функцій активації, H – число прихованих станів, D – розмірність ознак та Y – число міток класів. Під час використання алгоритму розповсюдження довіри для послідовності X довжиною T обчислювальна складність складатиме $O(TYH^2)$.

Кількість проміжків може бути неоднаковою, в залежності від конфігурації моделі, тому для зручності та подальшого позначення введемо індекс для їх опису $X^1 = \{x^1_1, \ldots , x^1_T\}$ та оператор $r(x^1_t)$, що створює посилання, яке повертає групу спостережень попереднього проміжку. Для $I = 1$ $r(x^1_t) = x_t$.

Для узагальнення послідовності X^1 на наступний проміжок X^{1+1}, використовується алгоритм, описаний в роботі [4] з модифікованою метрикою подібності (рис. 1).

Таким чином, $G=(V,E,W)$ – неорієнтований граф з множиною вершин V, ребер E і ваг W, які визначають схожість між двома вершинами.

Алгоритми об'єднує $r(x^{I+1}{}_s)$ та $r(x^{1+1}{}_t)$, якщо подібність між групами менше, ніж мінімальне внутрішнє відмінність, яка визначена наступним чином:

$$M\ Int(C_s, C_t) = \max(Int(C_s) + \tau(C_s), Int(C_t) + \tau(C_t)).$$

У цій формулі Int – внутрішня відмінність, що визначається за формулою: $Int(C) = \max\limits_{(s,t)\in mst(C,E)} M_{s,t}$, де mst – мінімальне остовне дерево, а – порогова функція.

У свою чергу, метрику подібності, під час розв'язання даної задачі, вираховуємо наступним чином:

$$M_{s,t} = \sum_{y,h} \left| p(h_s = h' | y, X; W) - p(h_t = h' | y, X; W) \right|.$$

Згідно з [4] обчислювальна складність алгоритму становить $O(T\ logT)$, де T – довжина послідовності.

Умовна ймовірність, з урахуванням змінного числа проміжків, може бути визначена як $\prod_{I=1}^{L} p(y | X^I; W^I)$.

Оскільки W^I – вектор параметрів для поточного проміжку, то вектор параметрів для всієї моделі $W = \{W^1, \ldots, W^I\}$.

Представимо навчальну вибірку $D = (X_i, y_i)$, $X_i \in M^{D\times T_i}$, $y_i \in Y$, $i = 1 \ldots N$. Типовий підхід до пошуку оптимального рішення W^* полягає в мінімізації функції втрат:

$$\min_W L(W) = \frac{1}{2\sigma^2} \| W \| - \sum_{i=1}^{N} \log p(y_i | X_i; W).$$

Для даної моделі такий підхід не може бути застосований, тому що для отримання узагальнення послідовності на наступному проміжку X^{1+1} потрібно знати $p(h^1 | y, X^1; W^1)$.

Відповідним підходом буде послідовна оптимізація для кожного окремого проміжку I [5]:

$$\min_W L(W) = \frac{1}{2\sigma^2} \| W \| - \sum_{i=1}^{N} \log p(y_i | X_i; W).$$

Дана задача також може бути вирішена за допомогою квазіньютонівських методів, зокрема L-BFGS [6].

Таким чином, наступні методи були успішно застосовані для вирішення аналогічного класу задач [1].

Частинні похідні за параметром W^1 для навчальної вибірки D вираховуються як:

$$\frac{\partial \log p(y_i | X_i^I; W^I)}{\partial W^I} = \sum_{h^I} p(h^1 | y_i, X_i^I; W^I) \frac{\partial F}{\partial W^1} - $$
$$- \sum_{y',h^I} p(y' | h^I, X_i^I; W^I) \frac{\partial F}{\partial W^1}.$$

Запис частинних похідних $\frac{\partial F}{\partial W^1}$ відносно $w_{y,h}$ та $w_{y,h,h}$ аналогічний тому, що представлений у роботі [1]:

$$\frac{\partial f^2}{\partial w^I{}_{y,h}} = \sum_t I(y = y') I(h_t^I = h'),$$

$$\frac{\partial f^3}{\partial w^I{}_{y,h,h}} = \sum_t I(y = y') I(h_t^I = h') I(h_t^I = h''),$$

$$\frac{\partial f^1}{\partial w^I{}_{g,h}} = \sum_t I(h_t^I = h') \psi_g (X^I, t; W^I),$$

$$\frac{\partial f^1}{\partial w^I{}_{g,d}} = \sum_t w_{g,d}^I \frac{1}{|r(X_t^I|} \sum_x g(\partial w^I{}_{g,d} x')(1 - g(\partial w_{g,d}^I X').$$

Рисунки 2 та 3 демонструють процедуру навчання та перевірки моделі. Під час навчання для кожного проміжку I вираховується W^{*I} та створюється узагальнення X^{I+I} для кожного попереднього проміжку.

Під час перевірки проводиться підсумовування $\log p(y_i | X_I^I; W^I)$ та визначається оптимальна мітка класу.

Та якщо X^{I+I}=X^I, то процедура навчання та перевірки припиняється для X^I. Таким чином, X^{I+I} буде завжди коротша, ніж X^I.

Обчислювальна складність всієї моделі складається з $O(TYH^2)$ та $O(TlogT)$. Для L проміжків

```
INPUT: Граф G
OUTPUT: C = {c1, …, ck}
C =V, ct = c(xt^I+1) = {xt^I}, ∀t
O = SORT(E,W)
FOR q=1 TO |O|:
    s,t =Oq
    IF cs!= ct && Ms,t ≤ MInt(Cs, Ct):
        C = JOIN (Cs, Ct)
```

Рис. 1. Узагальнення послідовності шляхом використання алгоритму угруповання

```
INPUT: Навчальна вибірка D
OUTPUT: Оптимальне рішення w*
FOR I=1 TO L:
    W* = argmin L(w^I)
    FOREACH xi ∈ D
        xi^I+1 = JOIN (xi^I, w*I)
```

Рис. 2. Алгоритми навчання моделі

```
INPUT: Послідовність x,
       Оптимальне рішення W*
OUTPUT: Мітка класу y*
p(y|x; w*) = 0
FOR I=1 TO L:
    log p(y|x; w*)+= p(y|x^I; w*^I)
    x^I = JOIN (x_i^I, w*^I)
y* = argmax log p(y|x; w*)
```

Рис. 3. Алгоритм перевірки

вона буде складати: $O(LTYH^2 + LTlogT)$. Оскільки T – довжина кожного проміжку, то вона буде зменшуватися зі збільшенням порядку проміжків, тому що X^{I+1} буде завжди коротша, ніж X^I. Отже, обчислювальна складність моделі зростає сублінійно відносно числа її проміжків.

Висновки. Сукупність засобів, методів і правил взаємодії (управління, контролю тощо) між елементами системи на основі динамічних алфавітно-цифрових рядів має ряд вимог, таких як робота в режимі реального часу, можливість створення спеціальних баз запису даних, здатність реєстрації зміни конфігурації генератора алфавітно-цифрових рядів із заданою частотою, збереження високої точності розпізнавання тощо.

Таким чином, у результаті проведеного дослідження показано, що існуючі методи розпізнавання динамічних алфавітно-цифрових рядів володіють функціональними обмеженнями і не можуть у повному обсязі реалізувати практично необхідні вимоги. Тому була запропонована і описана на теоретичному рівні модифікація методу умовних випадкових полів із прихованими станами. Дана модифікація методу дозволяє усунути ряд обмежень моделі та підвищити її дискримінативні здібності.

У ході роботи показано, що модифікація методу умовних випадкових полів із прихованими станами алгоритмічно може бути реалізована і має сублінійну часову складність.

Список літератури:

1. Vail Douglas L. Conditional Random Fields for Activity Recognition. PhD Thesis, Carnegie Mellon University, United States of America, CMU-CS-08-119, April, 2008.
2. Sutton Charles, McCallum Andrew. An Introduction to Conditional Random Fields. Foundations and Trends in Machine Learning .2012. Vol. 4, No. 4 (2011). P. 267–373. DOI: 10.1561/2200000013/
3. Do T., Artieres T. Neural conditional random fields. International Conference on Artificial Intelligence and Statistics (AI-STATS). P. 177–184, 2010.
4. Felzenszwalb P.F., Huttenlocher D.P. Efficient graph-based image segmentation. International journal of computer vision. Vol. 59, No. 2. September 2004. USA/
5. Hartline J. R. K. Incremental Optimization. PhD Thesis. Faculty of the Graduate School, Cornell University. 2008.
6. Nocedal J., Wright S. J. Numerical Optimization. Springer, USA. 1999.

АДАПТИВНЫЙ МЕТОД РАСПОЗНАВАНИЯ ДИНАМИЧЕСКОГО АЛФАВИТНО-ЦИФРОВОГО РЯДА

В статье проведен анализ актуальности и методов распознавания динамических алфавитно-цифровых рядов. Предложена модификация метода условных случайных полей со скрытыми состояниями. Она позволяет повысить дискриминативные способности данной модели. Проведено исследование модели на алгоритмическом уровне. Установлено, что данная модификация алгоритмически реализуема и обладает сублинейной временной сложностью.

Ключевые слова: *динамический алфавитно-цифровой ряд, распознавание образов, метод условных случайных полей со скрытыми состояниями, генеративная вероятностная модель, дискриминативная вероятностная модель.*

ADAPTIVE METHOD OF RECOGNITION FOR A DYNAMIC ALPHANUMERIC SERIES

The scientific article is analyzed for relevance and methods of recognition for dynamic alphanumeric series. A modification was proposed for the method of conditional random fields with hidden conditions. It allows increasing the discriminative abilities of this model. The research of the model was conducted at the algorithmic level. This modification is implemented by the algorithm and has a sublinear time complexity.

Key words: *dynamic alphanumeric series, pattern recognition, the method of conditional random fields with hidden conditions, generative probability model, discriminative probability model.*