MSC 90C35, 90C27 DOI: 10.14529/mmp230404

EFFECTIVE PRACTICES OF USING SPATIAL MODELS IN DOCUMENT IMAGE CLASSIFICATION

O.A. Slavin[1,2]***, I.M. Janiszewski***[1]
[1]Federal Research Center "Computer Science and Control" RAS, Moscow, Russian Federation
[2]LLC "Smart Engines Service", Moscow, Russian Federation
E-mails: oslavin@isa.ru, yanishevsky@isa.ru

This paper presents a new approach to modelling the structure of document images for classification tasks. Each of the document images is considered as a realization of a stochastic point process. Estimates of the properties of the point process are used to describe the document structure. The main objective of this paper is to determine the type of a new document using a nonparametric classification method. A method of classification of functional properties of point processes based on the concept of statistical depth is proposed. Practical issues of experimentation are considered. Modeling on real data showed the effectiveness of the proposed approach.

Keywords: documents with flexible structure; classification; spatial point process; reproducible point patterns; depth; DD-plot; α-procedure.

Introduction

Recently, there has been a growing need for efficient technologies in the field of automated document image processing (DIP). Document image classification is a crucial component of the DIP system [1–3]. When acquiring a document image, it is essential to categorize it into one or more predefined document types.

This problem can be approached in two ways: as an image classification task or as a text classification task. The former involves identifying patterns within the image pixels and evaluating structural features. Document classifiers relying on textual content utilize optical character recognition (OCR) techniques to extract text from the document image, which may introduce potential OCR errors.

The key aspects of document classification are as follows:

- document feature representation;
- models of document types;
- classification algorithms;
- training mechanisms.

Document classification is instrumental in sorting document image streams. These documents can be either single or multi-paged. For this study, we specifically examine single-page documents. We consider document streams of three types:

- stream of documents of a single, pre-defined class;
- stream of documents spanning several pre-defined classes;
- stream of documents spanning both pre-defined and unknown classes.

After sorting, each document is assigned to one of the known classes, or it is deemed unclassifiable. The number of document types in the stream is arbitrary; we focused on streams with a maximum of 10 classes.

1. Model Description

Business documents are commonly categorized into those with rigid and flexible structures [4]. In a document with a rigid structure, the layout and static text remain fixed. In a document with a flexible structure, they undergo various alterations. It is important to note that the positioning of words in documents with a flexible structure depends on a combination of several random factors:

- font and font size changes;
- adjustments in line spacing;
- text wrapping to the next line;
- text wrapping to the next page;
- removal of words in static text;
- substitution of words in static text.

Let us represent the business document D with flexible structure as a finite set of special text key points $W = \{T, B\}$ [4], where

- T is the core of a special text key point, which is represented by a sequence of characters $s_1, s_2, \ldots, s_n$, of some alphabet;
- B denotes the boundary, comprising coordinates of a quadrilateral that bounds the image of the special text key point, along with quadrilaterals that enclose the images of each character. These coordinates are normalized by the height and width of the page.

The tuple $\{T, B\}$ defines the descriptor of the special text key point. The special text key point detector is an OCR procedure.

The feature point is a counterpart to the special text key point. Common examples of image feature points are corners, endpoints of line segments, and other topological features of an image's morphological skeleton. In document recognition, feature points are used to classify and localize documents, or their sections, by comparing them to reference documents. The approach involves detecting all feature points, rectifying them (e.g., via RANSAC), and utilizing sets of feature points known as constellations.

Similarly, special text key points and constellations of points can be employed for document classification. Within a constellation, for each pair of special text key points (A, B), the following relation can be defined: $A \otimes B$. Other relations include:

- $\omega \in \varepsilon$ – point ω belongs to the fragment ε, where the words are linearly ordered;
- $\omega_1 < \omega_2$ – point ω_1 is positioned "prior to" point ω_2;
- $\omega_1 < \omega_2$ – point ω_1 is positioned "above" point ω_2.

The following metrics can be used:

- $d_1(\omega_1, \omega_2)$ counts the number of points within the interval between ω_1 and ω_2,
- $d_2(\omega_1, \omega_2)$ measures the sum of widths of points located between ω_1 и ω_2,
- $d_x(\omega_1, \omega_2)$, $d_y(\omega_1, \omega_2)$ – the Euclidean distance between the projections of the boundaries of ω_1 and ω_2.

Using the described operations and metrics, constellations can be represented as chains – sequences of points $\omega_1, \ldots, \omega_n$, where each pair of points is ordered $\omega_i < \omega_{i+1}$. For certain pairs of points (ω_i, ω_j) only one or several relations are defined. A straightforward example of a chain is a sequence of points within a single text line, with the simplest case being a shingle. Another instance is a sequence of points arranged vertically as $\omega_i \vee \omega_{i+1}$. Document image classification involves linking all points in the constellation. This entails attempting to map constellation points to the recognized words based on computed Levenshtein distances. During this process, all specified relations between chain points are validated. For both cases of simple chains, training entails constructing chains with a known structure using the fewest possible points. This requires a labeled training dataset containing documents from various classes. Training for chains in their general form is computationally demanding.

We will use a spatial point process as the mathematical representation for a set of coordinates. Let X denote the point process. Each coordinate of a special text key point will be viewed as an event. The collection of these coordinates within a document forms the realization of a point process or the point pattern. It is important to highlight the difference between the theoretical model, termed the point process, and its realization, a deterministic arrangement known as a point pattern (denoted by $\mathcal{X}$). When modelling documents with flexible structures, we are dealing with numerous realizations (point patterns) which can be regarded as realizations of the same point process. This collective ensemble of realizations will be referred to as replicated point patterns.

We utilize summary functions to capture statistical patterns in event arrangements [5, 6]. The functions employed include:

- G-function: Describes the nearest-neighbor distance distribution

$$\widehat{G}(r) = \frac{1}{N\left(W_{\ominus r}\right)} \sum_{[x:d(x)]} I_{W_{\ominus r}}(x) I\left(0 < d(x) \leq r\right),$$

  where $N\left(W_{\ominus r}\right)$ represents the number of observed events within the window $W_{\ominus r}$ (see Fig. 1), $d(x)$ is the Euclidean distance to the nearest neighbor, and I is the indicator function.

- F-function: Defines the distribution of distances from any randomly selected point to the nearest event

$$\widehat{F}(r) = \frac{\nu\left(\bigcup_{x \in X} b(x,\ r) \cap W_{\ominus r}\right)}{\nu\left(W_{\ominus r}\right)},$$

  where $b(x,\ r)$ is the disc with center x and radius r, and $\nu(A)$ denotes the volume of window A.

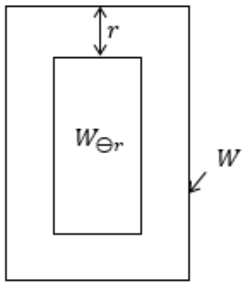

Fig. 1. Observation window W and reduced window $W_{\ominus r}$

- J-function:

$$\widehat{J}(r) = \frac{(1 - G(r))}{(1 - F(r))}, \widehat{F}(r) < 1.$$

- K-function: Determines the cumulative average number of data points lying within a distance r of a typical data point

$$\widehat{K}(r) = \frac{\nu(W)}{N(W)\,(N(W) - 1)} \sum_{x_1, x_2 \in X} I\,(0 < \|x_1 - x_2\| \leq r),$$

  where $N(W)$ stands for the observed events in window W, $\nu(W)$ represents the area of window W, and $\|\cdot\|$ signifies the Euclidean norm.

- L-function: A transformation of the K-function

$$\widehat{L}(r) = \sqrt{\frac{\widehat{K}(r)}{\pi}}.$$

Our approach assumes that, by employing these functions, the original problem can be redefined as a functional data classification problem [7, 8].

2. Studentized Permutation Test

Before diving into the classification task, it is prudent to confirm that the differences between the groups hold statistical significance. The null hypothesis suggests that there are no differences between the groups, implying that observed point patterns, regardless of their group affiliation, are independent and identically distributed random point patterns. The alternative hypothesis contends that the point patterns within the group are independent and identically distributed.

Consider g groups of reproducible point patterns, each comprising $n_1, \ldots, n_g$ patterns respectively. Let n_{ij} denote the count of events in the j-th point pattern of the i-th group, with $w_{ij} = \frac{n_{ij}}{n_i}$, $n_i = \sum\limits_{j=1}^{m_i} n_{ij}$ and $n = \sum\limits_{i=1}^{g} n_i$. For each group, we define the averaged function:

$$\overline{K}_i(r) = \sum_{j=1}^{m_i} w_{ij} \widehat{K}_{ij}(r), i = 1, \ldots, g.$$

Additionally, let us introduce the generalized averaged function:

$$\overline{K}(r) = \frac{1}{n}\sum_{i=1}^{g} n_i \overline{K}_i(r).$$

As a criterion for testing alternative hypotheses, we will employ the test proposed in [9],

$$T = \sum_{1\le i<j\le g} \int_0^{r_1} \frac{\left(\overline{K}_i(r) - \overline{K}_j(r)\right)^2}{\frac{1}{n_i}s_i^2(r) + \frac{1}{n_j}s_j^2(r)} dr,$$

where $s_i^2(r) = \frac{1}{n_i - 1}\sum\limits_{j=1}^{m_i}\left(\widehat{K}_{ij}(r) - \overline{K}_i(r)\right)^2$ represents the presumed within-group variances of estimates for distance r.

Note that instead of the K-function, one can employ any other summary function, such as G-,F-, L-, or J-function.

3. Functional Data Classification

In this section, we will describe a classification approach based on the concept of data depth. Data depth function assesses the proximity of an object to an implicitly defined class center. Let $\mathbf{y} \in R^d$, Y represent a random vector and its sample $\{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$ be from R^d, allowing us to estimate the vector's distribution. Examples of depth functions:

- the Mahalanobis depth [10] is defined as

$$D^{Mah}(\mathbf{y}|Y) = \left(1 + (\mathbf{y} - \mu_Y)^T \Sigma_Y^{-1} (\mathbf{y} - \mu_Y)\right)^{-1},$$

  where μ_Y determines the location of Y, and Σ_Y represents the spread of Y,

- the affine invariant spatial depth [11] is defined as

$$D^{Spt}(\mathbf{y}|Y) = 1 - \left\| E_Y\left[\upsilon\left(\Sigma_Y^{-\frac{1}{2}}(\mathbf{y} - Y)\right)\right]\right\|,$$

  where $\|\cdot\|$ represents the Euclidean norm, $\upsilon(\mathbf{w}) = \|\mathbf{w}\|^{-1}\mathbf{w}$ for $\mathbf{w} \neq \mathbf{0}$ and $\upsilon(\mathbf{0}) = \mathbf{0}$, Σ_Y is the covariance matrix of Y,

- the projection depth [12] is given by

$$D^{Prj}(\mathbf{y}|Y) = \inf_{\mathbf{u}\in S^{d-1}} \left(1 + O^{Prj}(\mathbf{y}|Y, \mathbf{u})\right)^{-1}$$

  with $O^{Prj}(\mathbf{y}|Y, \mathbf{u}) = \frac{\left|\mathbf{y}'\mathbf{u} - m\left(Y'\mathbf{u}\right)\right|}{\mathrm{MAD}\left(Y'\mathbf{u}\right)}$, where m denotes the univariate median and MAD the median absolute deviation from the median.

The depth function adheres to the following criteria

- affine invariance;
- upper semicontinuity;
- quasi-concavity with respect to its first argument;
- approaches zero as the first argument tends towards infinity.

Mosler and Mozharovskyi [13] proposed a two-stage procedure for transforming original functional data into a low-dimensional space, followed by classification. They suggested mapping the functional data into a finite-dimensional location-slope space, treating each observation as a vector of integrals representing its levels (location L) and first derivatives (slope S) over L and S subintervals of equal size. The data in the (L, S)-space are then transformed into a depth-depth plot (DD-plot [14]), using a multivariate function, resulting in a low-dimensional unit cube. Finally, observations from different classes are separated using a projectively invariant procedure known as the alpha-procedure. Typically, separation in the depth space can be achieved using established methods like linear discriminant functions, nearest neighbor classifier, and so forth.

4. Experiment

We conducted an experiment using own custom dataset, which included 1941 images of bank documents categorized into eight classes.

Table 1

Dataset description

Class	1	2	3	4	5	6	7	8
Quantity	149	59	109	34	37	1193	200	160

Examples of point patterns for all document types are illustrated in Fig. 2.

In the initial phase of the experiment, property function estimates were computed for each document utilizing the **spatstat** package [15]. Estimates for the G-functions of point patterns are depicted in Fig. 3.

Subsequently, we verified that the distinctions between document groups held statistical significance. The obtained p-value of 0,001 (see Fig. 4) indicates the rejection of the null hypothesis in favor of the alternative.

In the second stage, the dataset was split into training and testing sets (50:50).

Next, we trained four classifiers implemented in the **ddalpha** package [16]: linear discriminant analysis (*LDA*), k-nearest-neighbors (*kNN*) classification, the maximum-depth (*maxDepth*) classification and $DD\alpha$-classifier. Comparative testing results are presented in Table 2.

The experimental results show that all considered classifiers demonstrate good classification accuracy, and *kNN* exhibits the best generalization performance on the available data.

Fig. 2. Examples of document point patterns

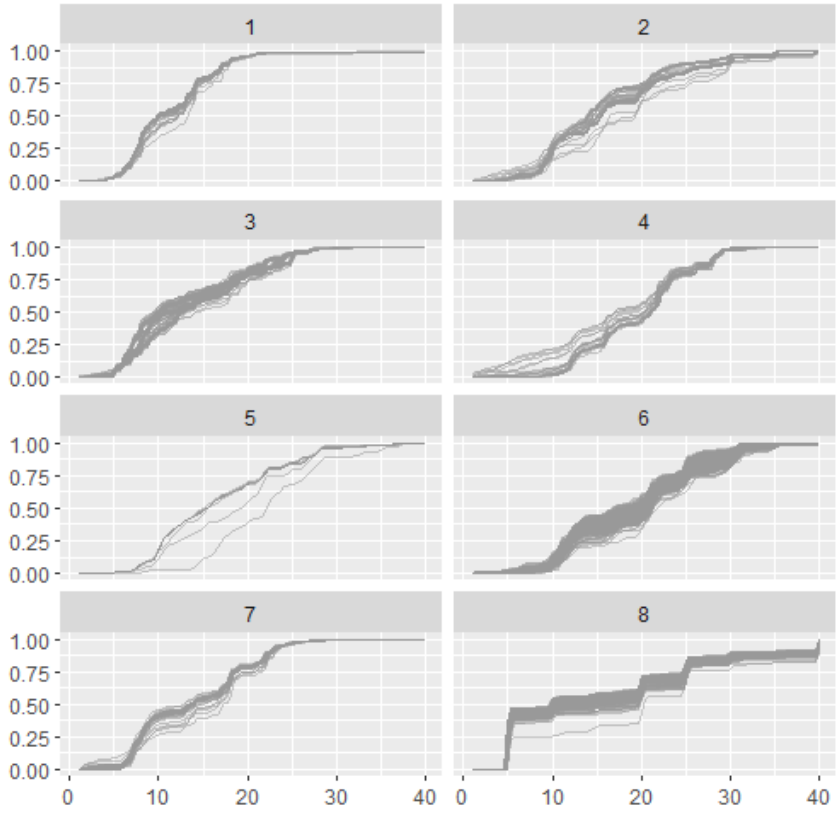

Fig. 3. G-function estimates

```
        Studentized permutation test for grouped point patterns
        item ~ type
        8 groups: 1, 2, 3, 4, 5, 6, 7, 8
        summary function: Gest, evaluated on r in [0, 150.301300018927]
        test statistic: T, 999 random permutations

data:  documents
T = 9271826, p-value = 0.001
alternative hypothesis: not the same G-function
```

Fig. 4. Hypothesis testing results

Table 2

Classifier testing results

Classifier\Function	K	L	G	F	J
DDα	0,9938	0,9917	0,9938	0,9835	0,9948
maxDepth	0,9928	0,9917	0,9938	0,9845	0,9948
kNN	0,9969	0,9958	0,9979	0,9958	0,9928
LDA	0,9907	0,9938	0,9928	0,9928	0,9938

Conclusion

This paper proposes a method for classifying images of documents with flexible structures using low-level structural information. We employ a spatial point process to model documents, representing each document as a point pattern. Collections of documents of the same type are represented as replicated point patterns. Experimental results confirm the effectiveness of the proposed approach.

Литература

1. Chen Nawei., Blostein D. A Survey of Document Image Classification: Problem Statement, Classifier Architecture and Performance Evaluation. *International Journal of Document Analysis and Recognition*, 2007, vol. 10, pp. 1–16. DOI: 10.1007/s10032-006-0020-2
2. Li Liu, Zhiyu Wang, Taorong Qiu, Qiu Chen, Yue Lu, Ching Y. Suen. Document Image Classification: Progress Over Two Decades. *Neurocomputing*, 2021, vol. 453, pp. 223–240. DOI: 10.1016/j.neucom.2021.04.114.
3. Gaceb D., Eglin V., Lebourgeois F. Classification of Business Documents for Real-Time Application. *Journal of Real-Time Image Processing*, 2014, vol. 9, no. 2, pp. 329–345. DOI: 10.1007/s11554-011-0227-4
4. Slavin O.A. Using Special Text Points in the Recognition of Documents. *Cyber-Physical Systems: Advances in Design and Modelling*, 2020, pp. 43–53.
5. Illian J., Penttinen A., Stoyan H., Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester, John Wiley and Sons, 2008.
6. Baddeley A., Rubak E., Turner R. *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton, London, New York, CRC press, 2015.
7. Pawlasová K., Dvořák J. Supervised Nonparametric Classification in the Context of Replicated Point Patterns. *Image Analysis and Stereology*, 2022, vol. 41, no. 2, pp. 57–109. DOI: 10.5566/ias.2652
8. Baíllo A., Cuevas A., Fraiman R. Classification Methods for Functional Data. *The Oxford Handbook of Functional Data Analysis*, Oxford, Oxford University Press, 2010, pp. 259–297.

9. Hahn U. A Studentized Permutation Test for the Comparison of Spatial Point Patterns. *Journal of the American Statistical Association*, 2012, vol. 107, pp. 754–764. DOI: 10.1080/01621459.2012.688463

10. Mahalanobis P.C. On the Generalized Distance in Statistics. *National Institute of Science of India*, 1936, vol. 2, no. 2, pp. 49–55.

11. Vardi Y., Cun-Hui Zhang. The Multivariate L_1-Median and Associated Data Depth. *Proceedings of the National Academy of Sciences*, 2000, vol. 97, no. 4, pp. 1423–1426. DOI: 10.1073/pnas.97.4.142

12. Zuo Yijun, Serfling R. General Notions of Statistical Depth Function. *Annals of statistics*, 2000, vol. 28, no. 2, pp. 461–482. DOI: 10.1214/aos/1016218226

13. Mosler K., Mozharovskyi P. Fast DD-Classification of Functional Data. *Statistical Papers 58*, 2017, vol. 4, pp. 1055–1089. DOI: 10.1007/s00362-015-0738-3

14. Li Jun, Cuesta-Albertos J.A., Liu R.Y. DD-Classifier: Nonparametric Classification Procedure Based on DD-plot. *Journal of the American Statistical Association*, 2012, vol. 107, no. 498, pp. 737–753. DOI: 10.1080/01621459.2012.688462

15. Baddeley A., Turner R. Spatstat: an R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, 2005, vol. 12, no. 6, pp. 1–42. DOI: 10.18637/jss.v012.i06

16. Pokotylo O., Mozharovskyi P., Dyckerhoff R. Depth and Depth-Based Classification with R-Package Ddalpha. *Journal of Statistical Software*, 2019, vol. 91, no. 5, pp. 1–46. DOI: 10.18637/jss.v091.i05

Received October 4, 2023

УДК 004.932.72'1 **DOI: 10.14529/mmp230404**

ЭФФЕКТИВНЫЕ ПРАКТИКИ ИСПОЛЬЗОВАНИЯ ПРОСТРАНСТВЕННОЙ МОДЕЛИ ПРИ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ ДОКУМЕНТОВ

***О.А. Славин*[1,2], *И.М. Янишевский*[1]**

[1]Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва, Российская Федерация

[2]ООО «Смарт Энджинс Сервис», г. Москва, Российская Федерация

В данной статье представлен новый подход к моделированию структуры изображений документов для задач классификации. Каждое из изображений документов рассматривается как реализация стохастического точечного процесса. Для описания структуры документа используются оценки свойств точечного процесса. Основная цель данной статьи – определить тип нового документа с помощью непараметрического метода классификации. Предлагается метод классификации функциональных свойств точечных процессов, основанный на понятии статистической глубины. Рассмотрены практические вопросы проведения эксперимента. Проведённое моделирование на реальных данных показали эффективность предложенного подхода.

Ключевые слова: гибкий документ; классификация; точечный процесс; воспроизводимые точечные паттерны; глубина; DD-диаграмма; α-процедура.

Олег Анатольевич Славин, доктор технических наук, доцент, главный научный сотрудник, Федеральный исследовательский центр «Информатика и управление» РАН (г. Москва, Российская Федерация); старший научный сотрудник-программист, ООО «Смарт Энджинс Сервис» (г. Москва, Российская Федерация), oslavin@isa.ru.

Игорь Михайлович Янишевский, кандидат физико-математических наук, старший научный сотрудник, Федеральный исследовательский центр «Информатика и управление» РАН (г. Москва, Российская Федерация), yanishevsky@isa.ru.

Поступила в редакцию 4 октября 2023 г.