Рисунок 4 – Области с зеленой растительностью

Заключение

В дальнейшем планируется оптимизация реализации алгоритма для сокращения времени анализа и повышения точности результатов. Также планируется создание системы, которая будет получать в реальном времени снимки с БПЛА, анализировать их и составлять карту поля с выделением проблемных участков.

Литература

1. Хорт, Д.О. Опыт и перспективы применения беспилотных летательных аппаратов в точном земледелии [Текст]/ Д.О. Хорт, Г.И. Личман, Р.А. Филиппов, А.И. Беленков// Нивы России -2016. –№5. – С. 62-67.

2. Sainz-Costa, N. Mapping Wide Row Crops with Video Sequences Acquired from a Tractor Moving at Treatment Speed [Текст]/ N. Sainz-Costa, A. Ribeiro, X.P. Burgos-Artizzu, M. Guijarro, G. Pajares// Sensors -2011. –№11. – С. 7095-7109.

3. Midtiby, H.S. Automatic Location of Crop Rows in UAV Images [Текст]/ H.S. Midtiby, J. Rasmussen// NJF Report -2014 –№5. – С. 22-25.

4. Jones, G. Modelling agronomic images for weed detection and comparison of crop/weed discrimination algorithm performance [Текст]/ G. Jones, Ch. Gée, F. Truchetet// Precision Agric -2009. –Volume 10, –Issue 1, – С. 1–15

5. OpenCV Documentation [Электронный ресурс] – http://opencv.org/documentation.html (дата обращения 12.02.2017).

А.Ю.Горчаков

ПРИМЕНЕНИЕ МЕТОДА НЕРАВНОМЕРНЫХ ПОКРЫТИЙ ДЛЯ БИНАРИЗАЦИИ ПРИЗНАКОВ В ЗАДАЧАХ КЛАССИФИКАЦИИ

(Вычислительный центр им. А.А. Дородницина Федерального исследовательского центра «Информатика и управление» Российской академии наук)

Задача бинарной классификации формулируется следующим образом. Пусть задано множество объектов X, множество меток Y = {0,1}, и существует целевая функция$y^*: X \to Y$, значения которой $y_i = y^*(x_i)$, известны только на конечном множестве объектов $X_1, \dots, X_n \in X$. Пары «объект-класс» (X_i, y_i) называются прецедентами.

Совокупность пар $(X_i, y_i)_{i=1}^{n}$ называется обучающей выборкой. Задача бинарной классификации заключается в том, чтобы по обучающей выборке научиться восстанавливать зависимость y^*, то есть построить решающую функцию $X \to Y$, которая бы приближала целевую функцию, причем не только на объектах обучающей выборки, но и на всем множестве X.

В случае если данные $X \in R$, некоторые из предлагаемых методов решения задачи [1],[2] предполагают бинаризацию этих данных.

Пусть $\varphi(x)$некоторый предикат, определенный на множестве объектов X, выделяет достаточно много объектов одного класса *C*, и практически не выделяет объекты другого класса. Введем обозначения:

P – число объектов класса *C*ввыборке

p – из них число объектов, для которых выполняется условие $\varphi(x) = 1$

N – число объектов не принадлежащих классу *C в* выборке

n – из них число объектов, для которых выполняется условие $\varphi(x) = 1$

Информативность предиката $\varphi(x)$ относительно класса $C \in Y$ по выборке $X^l = (X_i, y_i)_{i=1}^{l}$будем рассчитывать через статистическое определение информативности [2],[5]:

$$I_c(\varphi, X^l) = -ln\frac{C_P^p C_N^n}{C_{P+N}^{p+n}}, \text{где } 0 \le \mathrm{p} \le \mathrm{P}, 0 \le \mathrm{n} \le \mathrm{N}, (1.1)$$

где $C_m^k = \frac{m!}{k!(m-k)!}$ - биномиальные коэффициенты,

$0 \le \mathrm{k} \le \mathrm{m}$

Пусть $f: X \to R$ – числовой признак. Зонами значений признака *f*, будем называть предикаты вида:

$$\varphi(x) = [d \le f(x) \le d'], \qquad d < d' (1.2)$$

Требуется найти такие d и d', что $I_c(\varphi, X^l) \to max$.

Возьмем для примера выборку из примерно 100000 прецедентов, где $X_i \in [0,1]$, а множество меток Y = {0,1}.

1.Жадный алгоритм слияния зон [1], [2], [6] – разобьем интервал [0,1] на *n* равных подынтервалов и вычислим математическое ожидание y^*, на каждом из них и на всем интервале:

$$M = \sum_{i=1}^{l} y_i, (1.3)$$

$$M_j = \sum_{i=1}^{m} y_i, \text{ где } d_j \le f(i) < d_{j+1} \ (1.4)$$

Введем новый класс C', так что подынтервал принадлежит классу C' если $M_j > M$, и не принадлежит если $M_j \le M$. В качестве порогов возьмем d_i, такие что они лежат между всеми парами подынтервалов ровно один из которых принадлежит классу *C'*.

Таким образом начальное разбиение состоит их чередующихся зон «только *C'* – только не *C'*». Далее зоны укрупняются путем слияния троек соседних зон. Зоны сливаются до тех пор, пока информативность некоторой слитой зоны превышает информативность исходных зон, либо пока не будет получено заданное количество зон r.

Каждый раз выбирается та тройка, при слиянии которой достигается максимальный выигрыш информативности.

2.Метод неравномерных покрытий [6] (незначительная модификация метода, приведенного в [3], [4]) –

Предположим, что функция $f(x)$ удовлетворяет условию Липшица, то есть для любых x_1 и x_2 существует число $L\ < 0$ такое, что

$$|f(x_1) - f(x_2)| \le L\|x_1 - x_2\|, \|z\| = \left[\sum_{i=1}^{n}\left(z^{(i)}\right)^2\right]^{\frac{1}{2}} \quad (1.5)$$

и известны ее значения в точках $x_1, x_2, \dots x_k$ из (1.5) следует

$$f(x_k) - L\|x - x_k\| \le f(x) \le f(x_k) + L\|x - x_k\|. \quad (1.6)$$

Определим величину

$$F_k = max[f(x_1), f(x_2), \dots, f(x_k)] (1.7)$$

Найдем множество Δ_k такое, что на Δ_k имеет место

$$f(x) \le F_k + \varepsilon \quad (1.8)$$

Условие (1.7) выполнено для всех x, удовлетворяющих хотя бы одному из k условий

$$f(x_j) + L\|x - x_j\| \le F_k + \varepsilon, j = 1,2, \dots, k. \quad (1.9)$$

При каждом фиксированном j значения x удовлетворяющие (1.9), заполняют *n*-мерный шар V_j, границей которого является сфера

$$\|x - x_j\| = (F_k - f(x_j) + \varepsilon)/L = R_j (1.10)$$

с центром в точке x_j и с радиусом, равным R_j.

Центры шаров с наименьшими радиусами $R_{min} = \varepsilon/L$ располагаются в тех точках x_j , где $f(x_j) = F_k$. Шар (1.9) и сферу (1.10) будем в дальнейшем обозначать одной буквой V_j.

Величина F_k является решением задачи об отыскании глобального максимума функции *f(x)* на множестве $\Delta_k = \bigcup_{j=1}^{k} V_j$, так как максимальное значение функции *f*, удовлетворяющей (1.8) не превосходит на множестве Δ_k
более чем на ε величину F_k.

Если для некоторой последовательности точек $x_1, x_2, \dots x_k$получено Δ_k покрывающее допустимое множество, то тогда F_k есть решение исходной задачи. Способов получения последовательностей таких точек может быть множество. Один из них – разбиваем множество на n-мерные кубы равного размера (по аналогии с методом перебора по равномерной сетке), а далее считаем куб покрытым, если он целиком содержится в одной из n-мерных сфер $V_1, V_2, \ldots, V_k$.

Результаты работы жадного алгоритма слияния зон, при различных количествах подынтервалов (n):

Значение n	Кол-во вычислений I_c	Максимум I_c	d	d'
11	55	34.172	0.0	0.2
51	1275	38.065	0.0	0.14
101	5050	38.065	0.0	0.14
501	125250	38.785	0.0	0.156

Таб.1

Теперь посмотрим, как работает алгоритм неравномерных покрытий:
Зададим параметры метода
$\varepsilon = 1.0$ и $L = 100.0$

Значение n	Кол-во вычислений I_c	Максимум I_c	d	d'
11	16	33.562	0.2	1.0
51	13	33.653	0.24	1.0
101	12	34.400	0.22	0.99
501	25	28.606	0.25	0.992

Таб.2

В работе [6] отмечалась важность корректной оценки константы Липшица, в данном случае подберем её, как это предлагается в работе [3], начав с некоторого значения $L = L_0$, решать задачу с $2L_0, 4L_0$ и т.д. до тех пор, пока результат не будет отличаться от предыдущего значения не более чем на ε.
$\varepsilon = 1.0, n = 51$

Значение L	Кол-во вычислений I_c	Максимум I_c	d	d'
100	13	33.653	0.24	1.0
200	45	36.289	0.12	1.0
400	147	38.065	0.0	0.14
800	538	38.065	0.0	0.14

Таб.3

Из рис.5 и рис.6 видно, что для работы алгоритма неравномерных покрытий существенно важна оценка константы Липшица L. При заниженной константе метод неравномерных покрытий пропускает точку, в которой функция принимает максимальное значение, при завышенной – производится излишнее количество вычислений функции.

Сравнительный анализ алгоритмов показал, что метод перебора по равномерной сетке гарантированно находит максимум с заданной точностью, но требует произведения большого количества вычислений. «Жадный» алгоритм слияния зон обходится небольшим количеством вычислений, но нахождение максимума не гарантируется.

Метод неравномерных покрытий, по вычислительной сложности аналогичен алгоритму «жадного» слияния зон и качеству нахождения максимума аналогичен методу перебора по равномерной сетке. Причем, в случае корректной оценки константы Липшица, метод неравномерных покрытий гарантированно находит значение глобального максимума с заданной точностью.

Литература

[1] Кузьмич Р.И., Гулакова Т.К., Масич И.С. Способы бинаризации разнотипных признаков в задачах классификации //Актуальные проблемы авиации и космонавтики, vol. 6, 2010, pp. 323-325.

[2] Воронцов К. В. Математические методы обучения по прецедентам (теория обучения машин), Москва, 2011.

[3] Евтушенко Ю. Г. Численный метод поиска глобального экстремума функций (перебор на неравномерной сетке) //Журнал вычислительной математики и математической физики, 1971, vol. 6. – pp.1390-1403.

[4] Evtushenko Y., Posypkin M. A deterministic approach to global box-constrained optimization //Optimization Letters, 2013, vol. 4, pp. 819-829.

[5] Dubner P. N. Statistical tests for feature selection in KORA recognition algorithms // Pattern Recognition and Image Analysis, 1994, Vol. 4, no. 4. p. 396.

[6] Gorchakov A. Y. Application of method nonuniform coverings for maximum information content of predicate search //International Journal of Open Information Technologies. 2017. T. 5. N. 2. – pp. 29-33.

Ю.О. Дюльдина, А.Р. Диязитдинова

МОДУЛЬ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ СОСТАВЛЕНИЯ КОММЕРЧЕСКИХ ПРЕДЛОЖЕНИЙ

(Поволжский государственный университет
телекоммуникаций и информатики)

Постановка задачи

В условиях обостряющейся конкуренции все большее значение приобретает стиль общения с клиентом. Любой бизнес – от крупных производственных предприятий до маленьких компаний – имеет одну и ту же конечную цель: продать как можно выгоднее результат своей работы, что приводит к ужесточению конкуренции на рынке, вынуждая владельцев бизнеса искать новые подходы к клиентам и инструменты продаж или совершенствовать имеющиеся. Современные исследования показывают, что наличие солидной базы лояльных клиентов является основным фактором устойчивости и процветания бизнеса компании.

Существует все меньше неавтоматизированных видов деятельности человека, одной их таких областей является сфера продаж. Пока невозможно посадить робота на совершение холодных звонков или послать его на встречу с кли-