

Жадные алгоритмы отбора признаков для решения задачи многоклассовой классификации

Е.Ф. Гончарова¹, А.В. Гайдель^{1,2}

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

²Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

Аннотация. Для улучшения результатов классификации изображений, полученных при ДЗЗ, на несколько классов, мы предлагаем применение двух жадных алгоритмов отбора признаков. Первый является алгоритмом жадного добавления признаков и основан на максимизации критерия дискриминантного анализа, второй – алгоритм жадного удаления признаков, который строится на основании линейной регрессионной модели. Главным преимуществом данных алгоритмов является то, что они вычисляют не просто эффективность каждого признака по отдельности, а рассчитывают информативность всего набора отобранных признаков. В связи с ограниченностью доступных для анализа изображений ДЗЗ, в данной работе применяется метод моделирования изображений, основанный на применении марковских случайных полей. В качестве оригинальных изображений используются изображения из базы данных UC Merced Land-Use, относящиеся к семи классам. Формирование признаков производилось с использованием программного обеспечения MaZda, позволяющего вычислять более 200 гистограммных и текстурных характеристик изображения. С помощью алгоритма жадного удаления признаков, основанного на построении линейной регрессионной модели, удалось понизить размерность признакового пространства с 218 до 15 признаков, причем доля верно классифицированных объектов составила 95%.

1. Введение

Задача многоклассовой классификации изображений является важнейшей задачей компьютерного зрения. Изображения могут характеризовать большим количеством числовых признаков. Так, например, программное обеспечение MaZda позволяет формировать более 200 разнообразных гистограммных и текстурных характеристик. Однако не все данные признаки одинаковым образом влияют на разделимость объектов, более того, рассмотрение всех возможных характеристик может значительным образом увеличить вычислительные затраты при проведении классификации.

В настоящее время широко исследуются методы отбора наиболее информативных признаков для решения задачи классификации. Так в работе [1] предлагается производить кластеризацию объектов в текущем признаковом пространстве последовательно добавлять признак, который максимизирует расстояния между центрами различных кластеров. В работе [2] рассматриваются методы отбора признаков для решения задачи классификации

гиперспектральных изображений поверхности Земли. Метод основан на сингулярном и QR-разложении матриц. Классификация производилась с помощью метода опорных векторов. Доля верно классифицированных объектов для различных наборов данных составила от 83% до 99%.

В данной работе для формирования признаков используется программное обеспечение MaZda, которое позволяет получать более 200 гистограммных и текстурных характеристик изображения [3].

Отбор признаков производится с помощью жадных алгоритмов добавления и удаления признаков, основанных на дискриминантном и регрессионном анализе, соответственно. Данные алгоритмы позволяют отбирать признаки, которые являются наиболее информативными для дальнейшей классификации объектов на заданное количество классов. Так как существует значительное ограничение на количество изображений, доступных для проведения исследования, в данной работе также рассматривается алгоритм моделирования изображений с помощью случайных марковских полей [4].

Изображения были получены из открытой базы данных UC Merced Land-Use Dataset, которая предоставляет изображения размерностью 256×256 отсчетов, относящиеся к семи различным классам. Для проверки полученных результатов была проведена классификация с помощью метода ближайшего соседа. Для того чтобы проверить, насколько качественно было проведено моделирование изображений, для них были рассчитаны векторы наилучших признаков, отобранных с помощью предложенных алгоритмов, которые, в свою очередь, были сравнены с теми же векторами для исходных реальных изображений.

2. Формирование признаков

Пусть изображение характеризуется своей матрицей яркости $I^{(M \times N)}$, где $M \times N$ – размер изображения в пикселах.

$$I(m, n) = \frac{R(m, n) + G(m, n) + B(m, n)}{3}, \quad m = \overline{1, M}, \quad n = \overline{1, N}, \quad (1)$$

где R, G, B – интенсивность красной, зеленой и синей составляющей отсчета с координатами (m, n) соответственно.

$I(m, n)$ принимает значения из промежутка от 0 до $I - 1$, где I – это максимально возможный уровень яркости изображения.

При формировании признаков мы вычисляем числовые характеристики изображения, которые впоследствии используются при проведении процедуры отбора признаков и последующей классификации. Для формирования признаков в данной работе использовалось программное обеспечение MaZda, с помощью которого можно рассчитать более 200 гистограммных и текстурных характеристик изображения.

Гистограмма строится на основании интенсивности отсчетов изображения, рассчитанных по формуле (1), без относительно к пространственной зависимости между отсчетами. Вычисляются следующие признаки: средняя интенсивность, дисперсия, коэффициент асимметрии и эксцесса, а также процентиля.

В следующую группу входят текстурные характеристики, рассчитанные по матрицам смежности уровня серого, которая вычисляется следующим образом:

$$P_{d_1, d_2}(i, j) = \{(m, n) \in \{1, 2, \dots, M\} \times \{1, 2, \dots, N\} \mid I(m, n) = i, I(m + d_1, n + d_2) = j\}, \quad i, j = \overline{0, L-1}.$$

Получаемые характеристики: второй угловой момент, контрастность, энтропия, корреляция, средние значения, дисперсия и энтропия для пяти различных расстояний по четырем направлениям $(i, 0), (0, i), (i, \pm i), i = \overline{1, 5}$.

Другая группа текстурных признаков строится на основании автокорреляционной функции, которая характеризует зависимость между отсчетами изображения. Признаки, рассчитанные на основании автокорреляционной функции, также вычисляются для пяти различных расстояний по четырем направлениям.

3. Методы отбора признаков

3.1. Задача отбора признаков

Отбор информативных признаков проводится с целью последующего решения задачи классификации. Пусть Ω – множество объектов, подлежащих распознаванию. С помощью разбиения $\Delta = \{\Omega_j\}_{j=0}^{L-1}$ множество Ω было разбито на L классов.

Для решения задачи классификации необходимо построить оператор $\tilde{\Phi}(x)$, который переводит x , вектор признаков объекта, $x \in \mathbb{R}^K$ (K – количество признаков), в его класс. $\tilde{\Phi}(x)$ должен быть максимально близок к идеальному оператору распознавания $\Phi(x)$, то есть такому, которому известна информация о реальном классе неизвестного объекта. Классификация входит в раздел обучения с учителем, поэтому для построения оператора $\tilde{\Phi}(x)$ необходимо использовать информацию, получаемую из обучающей выборки $U \subseteq \Omega$, для объектов которой класс известен.

Задача отбора признаков состоит в выборе некоторого подмножества признаков $Q \subseteq [1; K] \cap Z$, обеспечивающего минимальную ошибку классификации ε (3).

Для классификации векторов признаков был использован метод ближайшего соседа, согласно которому класс заданного вектора x определяется как класс его ближайшего соседа из обучающей выборки U . Расстояние между векторами признаков $\rho(x, y)$ рассчитывается с использованием метрики евклидова пространства:

$$\rho(x, y) = \sqrt{\sum_{i=1}^K (x_i - y_i)^2}, \quad x \in \mathbb{R}^K, y \in \mathbb{R}^K, \quad (2)$$

где K – количество признаков.

При этом вероятность ошибочного распознавания оценивается, как

$$\varepsilon = \frac{|\{x \in \tilde{U} \mid \Phi(x) \neq \tilde{\Phi}(x)\}|}{|\tilde{U}|}, \quad (3)$$

где \tilde{U} – контрольная выборка.

Контрольная выборка не должна пересекаться с обучающей выборкой, то есть $\tilde{U} \cap U = \emptyset$. Так как признаки могут быть разных масштабов, и по-разному влиять на классификацию, необходимо перейти к векторам с единичной дисперсией и нулевым математическим ожиданием, т.е. стандартизировать признаковое пространство. Для этого из каждого i -го признака необходимо вычесть оценку его математического ожидания и разделить на корень из оценки дисперсии.

3.2. Алгоритм жадного добавления признаков на основании дискриминантного анализа

В дискриминантном анализе критерий разделимости классов строится с помощью матриц рассеяния внутри класса и матриц рассеяния между классами [5]. Пусть x – это случайный вектор из признакового пространства, тогда нужно оценить, насколько сильно в этом пространстве отличаются векторы из разных классов.

Метод отбора признаков, основанный на дискриминантном анализе, был предложен в работе [5], в которой предлагается строить критерий разделимости двух классов с помощью матриц рассеяния внутри класса и матриц рассеяния между классами. В данной работе предлагается обобщение предложенного метода для построения критерия разделимости для нескольких классов. Тогда критерий разделимости будет строиться следующим образом, суммирование в знаменателе ведется по всем L классам:

$$J(\mathbf{Q}) = \frac{\text{tr } R}{\sum_{l=0}^{L-1} P(\Omega_l) \text{tr } R_l},$$

где \mathbf{Q} - текущий набор признаков;

R - корреляционная матрица смеси распределений;

R_l - корреляционная матрица внутри l -го класса;

$P(\Omega_l)$ - вероятность попадания объекта из класса Ω_l , для данной задачи $P(\Omega_l) = \frac{1}{2}$.

Для отбора признаков предлагается алгоритм жадного добавления признаков. Пусть изначально множество признаков пусто – $\mathbf{Q}_{(0)} = \emptyset$. На i -ом шаге рассматриваются все множества $\mathbf{Q}_{(i,j)} = \mathbf{Q}_{(i-1)} \cup \{j\}$, для которых вычисляется критерий $J_{i,j} = J(\mathbf{Q}_{(i,j)})$. Выбирается то из них, которое обеспечивает максимум критерия $J_{i,j}$, после чего действия повторяются.

$$\mathbf{Q}_{(i)} = \mathbf{Q}_{(i-1)} \cup \left\{ \arg \max_{j \in [1;K] \cap \mathbf{Z} \setminus \mathbf{Q}_{(i-1)}} J_{i,j} \right\} = \mathbf{Q}_{(i-1)} \cup \left\{ \arg \max_{j \in [1;K] \cap \mathbf{Z} \setminus \mathbf{Q}_{(i-1)}} J(\mathbf{Q}_{(i-1)} \cup \{j\}) \right\}.$$

Алгоритм останавливается, когда достигнуто необходимое количество признаков.

3.3. Алгоритм жадного удаления признаков на основании построения регрессионной модели

Предложенный алгоритм отбора является развитием алгоритма, представленного в работе [6], для случая многоклассовой классификации. Введем функцию $\Psi^l(x): \Xi \rightarrow [0;1] \cap \mathbf{Z}$, которая определяет, относится объект к классу l или нет. Функция $\Psi^l(x)$ задается следующим образом:

$$\Psi^l(x) = \begin{cases} 1, & y(x) = l, \\ 0, & y(x) \neq l. \end{cases}$$

Таким образом, $\Psi^l(x)$ – это зависимая переменная, которая находится в статистической зависимости от вектора признаков $x \in \Xi(\mathbf{Q})$. Чтобы оценить влияние признаков на классификацию объектов на L классов необходимо построить L уравнений линейной регрессии:

$$\Psi^l = X \theta^l + \varepsilon^l, \quad l = \overline{0, L-1},$$

где $\Psi^l = (\Psi^l_1 \quad \Psi^l_2 \quad \dots \quad \Psi^l_n)^T$ – выходной вектор, содержащий n значений;

X – матрица признаков размерности $n \times (|\mathbf{Q}| + 1)$, содержащая элементы x_{ij} – результаты i -го наблюдения за входным признаком x_j ;

$\theta^l = (\theta^l_0 \quad \theta^l_1 \quad \dots \quad \theta^l_{|\mathbf{Q}|})^T$ – вектор неизвестных коэффициентов регрессии;

$\varepsilon^l = (\varepsilon^l_1 \quad \varepsilon^l_2 \quad \dots \quad \varepsilon^l_n)^T$ – вектор ошибок, который содержит погрешности наблюдений.

Неизвестные коэффициенты, которые содержит вектор θ^l были оценены по обучающей выборке \mathbf{U} методом наименьших квадратов:

$$(\Psi^l - X \theta^l)^T (\Psi^l - X \theta^l) \rightarrow \min_{\theta^l}.$$

Таким образом, для каждого класса будет найдено L векторов θ^l , характеризующих коэффициенты в уравнении регрессии. Введем вектор $\hat{\theta} = (\hat{\theta}_1 \quad \hat{\theta}_2 \quad \dots \quad \hat{\theta}_{|\mathbf{Q}|})^T$, каждый элемент которого оценивается по формуле (4):

$$\hat{\theta}_i = \sum_{l=0}^{L-1} (\theta^l_i)^2, \quad i = \overline{1, |\mathbf{Q}|}. \quad (4)$$

Величина вклада каждого i -го признака оценивается в соответствии с коэффициентом $\hat{\theta}_i$, рассчитанным по формуле (4). Для отбора признаков предлагается алгоритм жадного исключения признаков. Пусть изначально множество содержит все возможные признаки $\mathbf{Q}_{(0)} = \mathbf{Q}$. На каждом j -ом шаге строятся L моделей линейной регрессии $\tilde{\Psi}_{(j)}^l = X_{(j)}^l \theta_{(j)}^l$ в очередном признаковом пространстве $\Xi(\mathbf{Q}_{(j)})$ и вычисляется вектор $\hat{\theta}_{(j)}$. Далее исключается признак с наименьшим по модулю элементом вектора $\hat{\theta}_{(j)}$, то есть $\mathbf{Q}_{(j+1)} = \mathbf{Q}_{(j)} \setminus \left\{ \arg \min_{k \in [1; K] \cap \mathbf{Z} \cap \mathbf{Q}_{(j)}} \left| \hat{\theta}_{(j)}(k) \right| \right\}$. Процесс продолжается, пока не будет достигнуто нужное количество признаков.

4. Моделирование изображений

4.1. Марковские случайные поля

Моделирование изображений предлагается производить на основании модели марковских случайных полей. Пусть $F = \{F_i \mid i \in S\}$ – многомерная случайная величина такая, что F_i является одномерной случайной величиной и принимает значения $\{F_1 = f_1, F_2 = f_2, \dots, F_n = f_n\}$, $S = \{1, 2, \dots, N\}$ – конечное множество индексов. Вероятность того, что случайная величина F_i примет значение f_i обозначается как $P(f_i)$. При введении случайной величины таким образом можно считать, что F – это случайное поле. Событие $(F_1 = f_1, \dots, F_n = f_n)$ есть конкретная реализация $f = (f_1, f_2, \dots, f_n)$ многомерной случайной величины F .

Введем систему окрестностей $N = \{N_i \mid \forall i \in S\}$, где N_i – множество элементов, соседствующих с i -ым узлом. Таким образом, предполагается, что те узлы, от которых могут зависеть локальные характеристики в узле i , входят в его окрестность.

Для марковских полей выполняется соотношение (5):

$$\forall f P(F_i = f_i \mid F_j = f_j, i \neq j) = P(F_i = f_i \mid F_j = f_j, j \in N_i). \quad (5)$$

Из данного соотношения следует, что марковские поля обладают свойством условной независимости [7]. Согласно (5) распределение F_i зависит от распределения отсчетов на окрестности N_i . Таким образом, если значения на окрестности определены, то значения F_j для $j \neq i$ и $j \notin N_i$ невлиятна F_i .

4.2. Моделирование изображений

Итак, предположим, что множество индексов S задает множество точек на плоскости. Дискретное изображение можно представить как реализацию двумерной случайной величины F в этих точках. Из свойства условной независимости следует, что значение интенсивности каждого отсчета изображения можно предсказать лишь на основании нескольких отсчетов, входящих в его окрестность.

Таким образом, моделирование изображений на основании марковских случайных полей происходит следующим образом. На каждом шаге k алгоритма для каждого i -го узла изображения $G_{(k)}(i)$, строится окрестность N_i , которая сравнивается с окрестностью того же узла на изображении-оригинале $G_{in}(i)$, на основании которого происходит моделирование. Значения узла изменяется согласно следующему правилу:

$$G_{(k+1)}(i) = G_{in} \left(\arg \min_{i_s \in S} \left(\rho \left(N_{i_s} \left(G_{in} \right), N_i \left(G_{(k)} \right) \right) \right) \right).$$

Метрика $\rho(x, y)$ определяется по формуле (2). Начальное изображение $G_{(0)}$ инициализируется белым шумом.

В данной работе используется каузальная окрестность размерности 5. Используемый в данной работе шаблон окрестности представлен на рисунке 1. Особенностью данной окрестности является то, что она содержит лишь те отсчеты изображения, которые предшествуют текущему обрабатываемому отсчету. Таким образом, значение каждого обрабатываемого отсчета (за исключением нескольких начальных) будет устанавливаться на основании уже обработанных отсчетов.

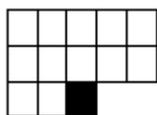


Рисунок 1. Пример каузальной окрестности размерности 5, где черным обозначен отсчет изображения, рассматриваемый в настоящий момент.

5. Эксперименты

5.1. Экспериментальные результаты отбора признаков

Для экспериментальной проверки предложенных методов была использована база данных UC Merced Land-Use Dataset, которая предоставляет изображения размерностью 256×256 отсчетов, относящиеся к различным классам. В каждом классе содержится по 100 изображений, полученных в ходе аэросъемки. В данной работе исследовались изображения, относящиеся к семи классам: поле, лес, пляж, постройки, поле для гольфа, заросли чапаралья и автомагистраль. На рисунке 2 приведены примеры изображений из репозитория.

При проведении экспериментов была применена пятикратная кросс-валидационная схема. С помощью программного обеспечения MaZda для каждого изображения было рассчитано 218 исходных признаков.

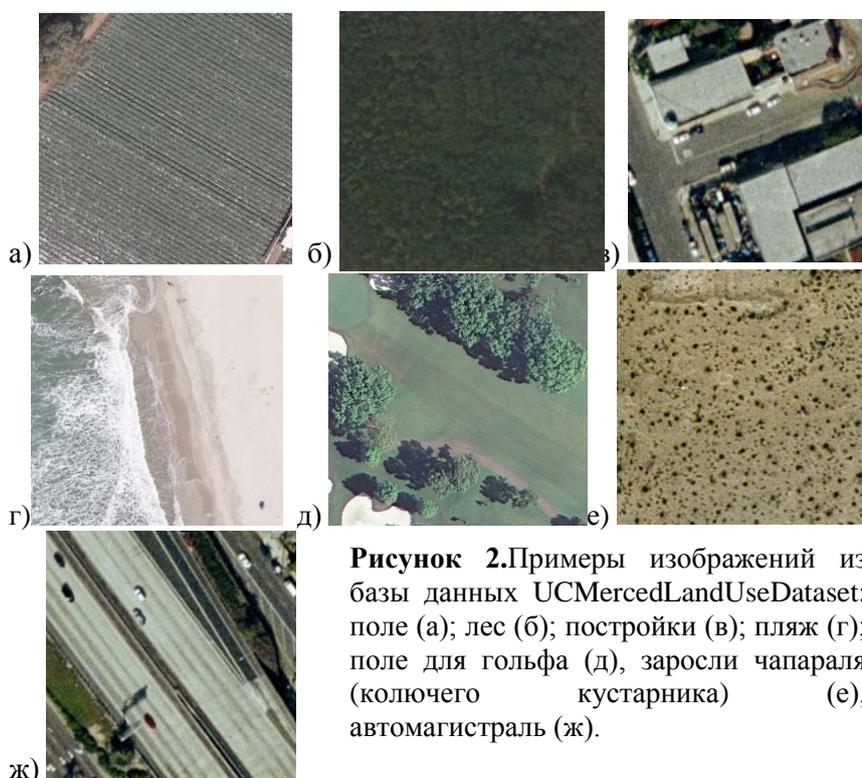


Рисунок 2. Примеры изображений из базы данных UCMercedLandUseDataset: поле (а); лес (б); постройки (в); пляж (г); поле для гольфа (д), заросли чапаралья (колючего кустарника) (е), автомагистраль (ж).

Результаты, полученные с помощью дискриминантного и регрессионного анализа, приведены в таблице 1.

Таблица 1. Признаки, отобранные с помощью дискриминантного и регрессионного анализа, отсортированные в порядке убывания значимости.

Дискриминантный анализ		Регрессионный анализ	
Порядковый номер признака	Название признака	Порядковый номер признака	Название признака
37	S11SumVarnс	96	S202DifEntrp
30	S01DifEntrp	74	S02DifEntrp
24	S01InvDfMom	85	S22DifEntrp
40	S11DifVarnс	107	S30DifEntrp
...
79	S22InvDfMom	171	S44Entropy
34	S11SumOfSqs	215	S55Entropy
32	S11Contrast	217	S55DifEntrp

Далее в таблицах 2 и 3 приведены группы признаков, отобранные с помощью двух методов, а также вероятность ошибочной классификации изображений в данных группах, рассчитанная по формуле (3).

Таблица 2. Группы наиболее информативных признаков, отобранные с помощью алгоритма жадного добавления.

K	Признаки	ε
3	37, 30, 24	0,74
4	37, 30, 24, 40	0,47
5	37, 30, 24, 40, 38	0,63
6	37, 30, 24, 40, 38, 2	0,63
...
47	37, 30, 24, 40, 38, 2, 55, 13, 42, 209, 44, ..., 91, 69, 80, 16, 118, 85	0,32
48	37, 30, 24, 40, 38, 2, 55, 13, 42, 209, 44, ..., 91, 69, 80, 16, 118, 85, 5	0,32

Таблица 3. Группы наиболее информативных признаков, отобранные с помощью алгоритма жадного удаления.

K	Признаки	ε
3	96, 74, 85	0,16
4	96, 74, 85, 107	0,32
5	96, 74, 85, 107, 140	0,42
6	96, 74, 85, 107, 140, 63	0,16
...
15	96, 74, 85, 107, 140, 63, 151, 129, 173, 118, 110, 154, 66, 99, 143	0,05
16	96, 74, 85, 107, 140, 63, 151, 129, 173, 118, 110, 154, 66, 99, 143, 138	0,05

Анализ полученных результатов показывает, что жадный алгоритм удаления признаков, основанный на построении линейной регрессионной модели, значительно лучше справился с поставленной задачей. В признаковом пространстве, состоящем из характеристик, отобранных с помощью данного алгоритма, минимальная вероятность ошибочной классификации изображений на семь классов составила 0,05. Данная ошибка была достигнута на наборе из 15 признаков из 218 исходных. Данная группа включает в себя текстурные признаки, сформированные для четырех расстояний 2, 3, 4 и 5. Алгоритм жадного добавления признаков, основанный на максимизации критерия дискриминантного анализа, показал худший результат для исследуемого набора данных, минимальная вероятность ошибочной классификации

составила 0,32, при этом, данная ошибка была достигнута на наборе признаков большей размерности (47 признаков). Стоит отметить, что вероятность ошибочной классификации, достигаемая на наборе признаков, содержащем в себе все 218 рассчитанных характеристик, составляет 0,37.

5.2. Экспериментальные результаты моделирования изображений

В качестве исходных изображений применялись изображения из исследуемого набора данных UC Merced Land-Use, переведенные в полутоновый формат. Результаты моделирования приведены на рисунке 3.

Для объективной проверки качества моделирования было проведено сравнение векторов признаков для смоделированных изображений и усредненных векторов для реальных изображений из базы данных UC Merced Land-Use. Векторы были построены из 15 лучших признаков, отобранных с помощью алгоритма жадного удаления признаков, построенного на основании линейной регрессии. Мера схожести $\xi(x, y)$, $x \in \mathbb{R}^K$, $y \in \mathbb{R}^K$ была введена как

$$\xi(x, y) = \left(\frac{1}{K} \sum_{k=1}^K (x_k - y_k)^2 \right)^{\frac{1}{2}}.$$

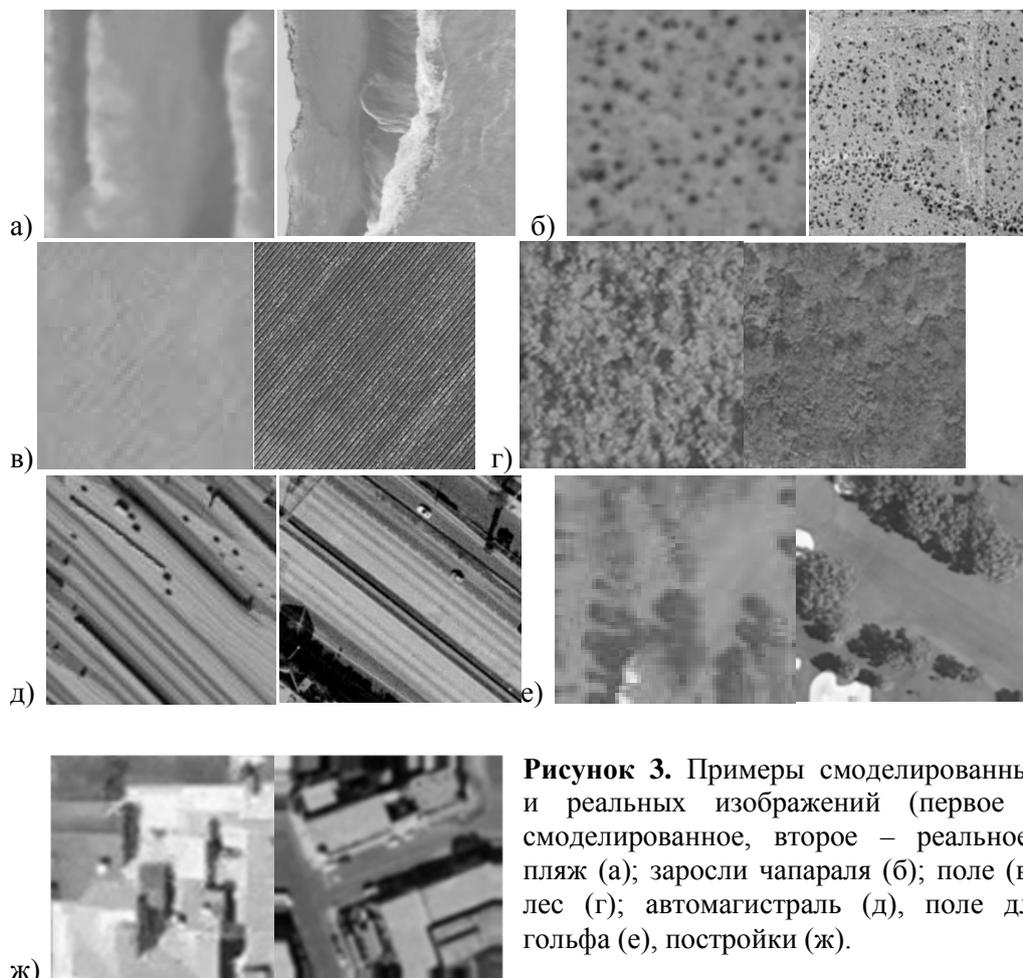


Рисунок 3. Примеры смоделированных и реальных изображений (первое – смоделированное, второе – реальное): пляж (а); заросли чапаралья (б); поле (в); лес (г); автомагистраль (д), поле для гольфа (е), постройки (ж).

Для изображений, смоделированных для каждого из семи классов, были получены значения $\xi(x, y)$, представленные в таблице 4.

Таблица 4. Значения меры схожести для смоделированных изображений, относящихся к одному из 7 классов.

Класс	$\xi(x, y)$
Пляж	0,15
Заросли чапаралья	0,08
Поле	0,14
Лес	0,09
Автомагистраль	0,09
Поле для гольфа	0,09
Постройки	0,13

Анализ полученных результатов показывает, что предложенный метод хорошо работает для изображений с ярко выраженной мелкой структурой, например, для изображений классов заросли чапаралья и лес. Изображения с крупными структурными элементами моделируются менее качественно. Для их моделирования необходимо применять окрестности больших размеров, что приводит к увеличению вычислительной сложности алгоритма. Для решения этого недостатка в работе [4] предлагается метод, основанный на построении гауссовых пирамид изображений, в результате чего крупномасштабные элементы на более низких уровнях пирамиды представляют данные крупные элементы в меньшем масштабе.

6. Заключение

Таким образом, в данной работе были исследованы два жадных алгоритма отбора наиболее информативных признаков для решения задачи многоклассовой классификации. Алгоритмы были экспериментально проверены на изображениях, полученных при ДЗЗ, из базы данных USMercedLand-Use, относящихся к семи различным классам.

Анализ результатов показал, что наименьшая вероятность ошибочной классификации достигается в признаковом пространстве, состоящем из 15 текстурных признаков, отобранных с помощью алгоритма жадного удаления признаков, построенного на основании модели линейной регрессии. Метод жадного добавления признаков показал худший результат по сравнению с предыдущим. Отметим, что исходное признаковое пространство состояло из 218 гистограммных и текстурных характеристик, сформированных с помощью программного обеспечения MazDa. Использование всех сформированных характеристик позволило верно классифицировать 67% изображений. В результате проведения процедуры отбора признаков удалось сократить размерность признакового пространства до 15 характеристик, при этом увеличить долю верно классифицированных объектов до 95%.

Для увеличения количества изображений, на которых можно обучать модели классификации и которые можно использовать для экспериментальной проверки предложенных алгоритмов, было проведено моделирование изображений с помощью марковских случайных полей. Сравнение смоделированных и реальных изображений показало, что метод применим для синтеза изображений с мелкими структурными элементами. Для синтеза изображений, содержащих более крупные элементы, необходимо усовершенствование рассмотренного метода, например, использование подхода, основанного на применении гауссовых пирамид.

7. Благодарности

Работа выполнена при поддержке Федерального агентства научных организаций (соглашение № 007-ГЗ/ЧЗ363/26).

8. Литература

- [1] Liu, C. A new feature selection method based on a validity index of feature subset / C. Liu, W. Wanga, Q. Zhao, X. Shen, M. Konan // *Pattern Recognition Letters*. – 2017. – Vol. 92. – P. 1-8.
- [2] Reshma, R. Dimensionality Reduction Using Band Selection Technique for Kernel Based Hyperspectral Image Classification / R. Reshma, V. Sowmya, K.P. Soman // *Procedia Computer Science*. – 2016. – Vol. 93. – P. 396-402.
- [3] Strzelecki, M. A software tool for automatic classification and segmentation of 2D/3D medical images / M. Strzelecki, P. Szczypinski, A. Materka, A. Klepaczko // *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. – 2013. – Vol. 702. – P. 137-140.
- [4] Wei, L.Y. Fast texture synthesis using tree-structured vector quantization / L.Y. Wei, M. Levoy // *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. – ACM Press/Addison-Wesley Publishing Co., 2000. – P. 479-488.
- [5] Гончарова, Е.Ф. Статистическое исследование факторов, влияющих на развитие сердечно-сосудистых заболеваний / Е.Ф. Гончарова, А.В. Гайдель, А.Г. Храмов // *Информационные технологии и нанотехнологии (ИТНТ-2016)*. – 2016. – С. 1020-1025.
- [6] Гончарова Е. Ф., Гайдель А. В. Методы отбора признаков для задач классификации изображений земной поверхности // *Информационные технологии и нанотехнологии (ИТНТ-2017)*. – 2017. – С. 535-540.
- [7] Winkler, G. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods* / G. Winkler–Springer-Verlag, 1995.

Greedy algorithms of feature selection for multiclass image classification

E.F.Goncharova¹, A.V.Gaidel^{1,2}

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

Abstract. To improve the performance of remote sensing images multiclass classification we propose two greedy algorithms of feature selection. The discriminant analysis criterion and regression coefficients are used as the measure of feature subset effectiveness in the first and second methods, respectively. The main benefit of the built algorithms is that they estimate not the individual criterion for each feature, but the general effectiveness of the feature subset. As there is a big limitation on the number of real remote sensing images, available for the analysis, we apply the Markov random model to enlarge the image dataset. As the pattern for image modelling, a random image belonging to one of the 7 classes from the UC Merced Land-Use dataset has been used. Feature shave been extracted with help of MaZda software. As the result, the largest fraction of correctly classified images accounts for 95%. Dimension of the initial feature space consisting of 218 features has been reduced to 15 features, using the greedy strategy of removing a feature, based on the linear regression model.

Keywords: feature selection, classification, remote sensing images, Markov model, greedy algorithms.