

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ВОССТАНОВЛЕНИЯ ЛЕСНОЙ РАСТИТЕЛЬНОСТИ ПО ФАКТОРАМ СРЕДЫ

Брижатая А.А.¹, Кислов Д.Е.²

¹*Ботанический сад-институт ДВО РАН, г. Владивосток*

²*Институт автоматизации и процессов управления ДВО РАН, г. Владивосток*

Выполнен сравнительный анализ методов решения задачи восстановления лесной растительности по материалам лесотаксационных описаний. Предложен способ, позволяющий оценить погрешности решения, обусловленные вынужденным использованием числовых методов для классификации объектов в качественных факторных пространствах. Для территории учебно-опытного хозяйства «Дальневосточный» (Приморский край, Уссурийский район) получены выводы о наиболее вероятных сукцессионных сменах в древесном ярусе.

Ключевые слова: восстановление лесной растительности, классификация, линейный дискриминантный анализ, метод *k*-ближайших соседей.

Лесные экосистемы играют важнейшую роль в развитии человечества и существования жизни на Земле в целом. Они занимают значительную часть земной поверхности, обеспечивая необходимыми для существования ресурсами множество жизненных форм, как непосредственно, являясь для одних организмов источником пищи и средой обитания, так и опосредовано, создавая благоприятные макроклиматические условия на планете для других. Несмотря на свою исключительную важность, лесная экосистема (при рассмотрении как в масштабах планеты, так в масштабах любой выделенной территории) представляет собой весьма малоизученный объект. На сегодняшний день многое неясно относительно того, какие факторы обеспечивают протекание сукцессионных процессов в лесных сообществах, каким образом влияют на структуру и динамику их развития изменения среды, связанные с антропогенными причинами или флуктуациями жизненно важных параметров и т.д. Изучение лесных экосистем представляет интерес не только с фундаментальной точки зрения, но и имеет ясно выраженную прикладную перспективу, ориентированную на управление процессами, связанными с использованием лесных ресурсов.

Одной из основных задач исследования лесных экосистем является установление взаимосвязей между лесной растительностью и факторами среды и их объективная (напри-

мер, количественная) характеристика (Василевич, 1969). Проблема восстановления лесной растительности, являющаяся предметом внимания настоящей работы, тесно связано с этой задачей. Восстановление исходной растительности подразумевает определение наиболее вероятного фитоценоза, способного произрастать в характерных для данного участка среды условиях. Если для какой-либо территории эта задача решается с высокой степенью точности – структура растительности восстанавливается с малой долей неверно определенных ценозов, то это указывает на наличие высокой степени взаимосвязи между условиями среды с одной стороны и структурой растительности с другой. Научные перспективы задачи восстановления растительности определяются, прежде всего, ее исключительной значимостью для разработки и верификации как аналитических (объясняющих) моделей динамики развития лесов, так и имитационных моделей, целью которых является по-возможности точное описание процессов, происходящих в лесных экосистемах. Знание структуры потенциально возможной растительности полезно не только для задач моделирования динамических процессов, но может использоваться и для выработки планов рационального лесопользования (Петропавловский, 2004).

С математической точки зрения задача восстановления лесной растительности является задачей классификации. Методология решения такого рода задач обеспечивается статистической теорией принятия решений и другими специальными разделами математической статистики (Рао, 1968). В зависимости от характера априорных сведений, наличия или отсутствия обучающей выборки или информации относительно количества классов, для решения задачи классификации применяются методы дискриминантного или кластерного анализа (Прикладная статистика..., 1989; Факторный..., 1989).

Восстановление фитоценозов по внешним факторам предполагает, что количе-

ство классов фиксировано и в распоряжении исследователя имеется обучающая выборка достаточного объема. Случаи, когда допускается принадлежность объекта к ранее не определенному классу, возникают, например, при выявлении нового ценоза, и здесь не рассматриваются. Целью предлагаемой работы является решение задачи восстановления лесных ценозов с использованием методов дискриминантного анализа.

Обзор математических методов классификации объектов

В настоящее время разработано множество различных методов, учитывающих как специфику конкретной решаемой задачи классификации, так и априорные сведения о принадлежности объектов заданным классам (Прикладная статистика..., 1989).

Классификационная процедура определяет отображение множества классифицируемых объектов во множество заданных классов на основе данных обучающей выборки. При этом, как правило, строится промежуточное отображение, сопоставляющее каждый классифицируемый элемент некоторым областям числового пространства (возможно многомерного), каждая из областей которого ассоциирована с определенным классом. Таким образом, если образ объекта принадлежит подмножеству, соответствующему заданному классу, то полагается, что и объект принадлежит этому классу. Такая интерпретация процедуры классификации данных является достаточно общей и соответствует практически любому методу дискриминантного анализа (Härdle, Hlavka, 2007).

Одним из распространенных алгоритмов является линейная дискриминантная классификация (*linear discriminate classifier, LDC*). Сущность метода заключается в определении набора классифицирующих функций вида $f_i(x) = u_i^T x + c_i$, $i = \overline{1, k}$, где k – число классов, зависящих от вектора факторов x в числовом пространстве известной размерности. Выбор массива коэффициентов $\{u_i, c_i\}$ осуществляется из условия минимума ошибочной классификации в предположении, что все внутриклассовые распределения нормальны и имеют равные ковариационные матрицы (Прикладная статистика..., 1989). Решающее правило метода состоит в том, чтобы отнести классифицируемый объект i -му классу, для которого f_i – максимальна.

Лесотаксационные данные не обладают свойством нормальности внутриклассовых распределений. В этом можно удостовериться, если рассмотреть графики однофакторной встречаемости видов (при нормальном распределении они должны иметь куполообразную форму), а также двухфакторных – когда линии уровня эмпирической оценки вероятности должны быть эллипсоидальными. Поскольку этого не наблюдается (визуальный тест приводит к отрицательному результату), то и пропадает необходимость проведения более строгой процедуры проверки нормальности, например, при помощи критерия χ^2 . Невыполнение фундаментальных положений для использования линейной дискриминантной классификации не является существенным барьером для его практического применения, тем более что он сохраняет высокую точность классификации и в этих случаях (Kuncheva, 2004). Однако, учитывая возможность сильных отклонений от нормального распределения, целесообразно рассмотреть альтернативу рассмотренному методу.

Такой альтернативой является метод k -ближайших соседей (Keller, 1986; Kuncheva, 2004). Данный метод относится к группе непараметрических методов. Его преимуществом является то, что он не требует никаких априорных сведений о характере внутриклассовых распределений данных. Он состоит в следующем: 1) в факторном пространстве вводится некоторая метрика; 2) в зависимости от объема обучающего массива выбирается целое число k ; 3) вокруг классифицируемой точки строится сфера (в заданной метрике) наименьшего радиуса, содержащая не менее k точек из обучающей выборки; 4) классифицируемая точка относится к той совокупности, к которой принадлежит большинство точек, попавших в построенную сферу.

Метод имеет различные модификации, заключающиеся в специфике метризации факторного пространства и выборе множества окрестностей точки в альтернативу сферам. Со статистической позиции назначение реализуемой методом процедуры – оценка условных плотностей вероятности в точке, которую необходимо классифицировать. Обобщением метода k -ближайших соседей является метод оценки плотности вероятности Парзена-Розенблатта (Kuncheva, 2004) с последующей процедурой классификации, например, на основе отношения правдоподобия (Pao, 1968).

В настоящей работе для решения задачи классификации выберем метрику

$$\rho(x, y) = \sum_{i=1}^s \frac{(x_i - y_i)^2}{n_i^2},$$

где n_i – количество градаций i -го фактора; s – количество учитываемых факторов, а в качестве окрестностей классифицируемых точек – сферы, связанные с этой метрикой. Такой выбор позволяет учесть различия в количествах градаций факторов.

Существенным препятствием для практического применения вышеописанных методов является качественный характер показателей факторных шкал. Примером является ситуация, когда вид подроста (нечисловой фактор) используется наряду с крутизной и высотой над уровнем моря для определения типа леса на данном участке местности. Наличие качественных показателей не позволяет осуществлять арифметические операции с факторными переменными и таким образом делает невозможным применение метода линейного дискриминантного анализа и метода k -ближайших соседей. Для последнего метода сложность заключается в невозможности метризации нечислового факторного пространства. Невольно напрашивается «тривиальное» решение указанной проблемы: взаимнооднозначно сопоставить каждой градации нечислового фактора некоторое натуральное число (здесь полагается, что мощность множества градаций качественной переменной конечна, что типично для рассматриваемых задач классификации; можно предположить, однако, что существуют задачи, в которых указанная мощность континуальна; в этих случаях взаимнооднозначного назначения натуральных чисел качественным градациям выполнить нельзя), заменив тем самым качественные показатели на количественные, и таким образом получить возможность беспрепятственно использовать вышеизложенные методы. Негативной стороной такого подхода является введение ложных представлений о соотношениях градаций факторов (например, изначально неупорядоченные градации наследуют от натуральных чисел отношение порядка). отождествление качественных показателей с натуральными числами «безопасно», пока с ними (числовыми образами качественных градаций) не производится арифметических операций и операций сравнения. Невозможность выполнять какие бы то ни было математические операции с полученными числовыми образами качественных переменных – очень жесткое требование, не по-

зволяющее применить ни один из представленных выше методов. Следовательно, необходимо каким-либо образом ослабить это требование, предложив ему альтернативу. Вполне логично рассуждать следующим образом. Если алгоритм решения (обозначим его K) задачи (в принципе произвольной, не обязательно задачи классификации) производит математические операции с числовыми образами качественных данных, то результат его применения зависит от примененного взаимнооднозначного соответствия, сопоставляющего качественным значениям действительные числа. Эта зависимость определяется привнесенной дополнительной (ложной) информацией, обусловленной преобразованием (обозначим его φ) качественных показателей в количественные, и вообще говоря, действием самого алгоритма K с полученными числовыми значениями. Требование того, чтобы результат выполнения алгоритма K не зависел от любого биективного (взаимнооднозначного) преобразования φ , обеспечит сведение к минимуму действия ложной информации на конечный результат и «обоснует» назначение качественным градациям числовых значений (в случае нескольких качественных показателей необходимо рассматривать совокупность преобразований $\{\varphi\}$). Обладающий таким свойством алгоритм K справедливо назвать φ -инвариантным.

Формально, наличие или отсутствие у алгоритма свойства φ -инвариантности не является серьезным препятствием его практического применения, однако, обладание этим свойством является более желательным, что отражает факт независимости конечного результата от способа преобразования «качество»–«количество». Нетрудно видеть, что представленные выше методы классификации не обладают свойством φ -инвариантности. Если, тем не менее, возникает необходимость классификации такими методами данных в факторном пространстве с нечисловыми «осями», следует иметь в виду, что результат классификации во многом будет определяться преобразованием, используемыми для назначения качественным градациям числовых значений. Таким образом, для алгоритма, не обладающего свойством φ -инвариантности, наблюдается неопределенность результата относительно выбора преобразования φ .

Рассмотрим альтернативу понятию φ -инвариантности, сузив класс допустимых

преобразований $\{\varphi\}$, представив его в виде суперпозиции фиксированного первого отображения (π), действующего из множества качественных показателей в некоторое известное подмножество X , например, натуральных чисел, и однопараметрического класса биекций $\{l_\alpha\}: X \rightarrow Y_\alpha, \alpha \in I$, где α и I – индекс параметризации и множество допустимых его значений соответственно. Таким образом, $\{\varphi\} = \{l_\alpha\} \circ \pi$. Допустим, что теоретически или экспериментально установлено, что результат действия метода остается неизменным для любого α из $S \subset I$. В зависимости от класса биективных преобразований $\{l_\alpha\}$ и множества S можно говорить о степени «эффективности» метода для решения задачи в нечисловых факторных пространствах. Предельным случаем такой меры "эффективности" является свойство φ -инвариантности алгоритма, когда результат его действия не зависит от выбора $\{l_\alpha\}$ и S .

Проиллюстрируем изложенное следующим примером. Предположим, что решается задача восстановления доминантов лесной растительности по трем геоморфологическим факторам: высоте над уровнем моря (в метрах), крутизне склона (в градусах) и экспозиции склона. Первые два фактора имеют числовую природу, однако последний содержит такие градации как "Пойма", "Равнина" и, таким образом, является качественным. Пусть число градаций качественного фактора конечно. Перенумеруем все возможные градации фактора и определим отображение π , назначив каждой градации её номер, т.е. π : (градация фактора) \rightarrow (номер градации фактора). В качестве класса биективных отображений $\{l_\alpha\}$ рассмотрим однопараметрическое семейство масштабирующих преобразований $\{l_\alpha: x \mapsto \alpha x\}$. Пусть, начиная с некоторого α^* , для любого $\alpha > \alpha^*$ (т.е. $S = \{\alpha: \alpha > \alpha^*\}$) наблюдается независимость результата выполнения алгоритма классификации (действующего в числовом факторном пространстве) от α . Такую ситуацию можно рассматривать как определенный произвол в назначении качественным градациям числовых значений: действительно, для любого назначаемого отображения $\varphi = l_\alpha \circ \pi, \alpha > \alpha^*$ итог выполнения алгоритма классификации одинаков, каков он и должен быть при оперировании качественными

показателями. Геометрически, при возрастании параметра α происходит разнесение данных вдоль оси – числового образа качественного фактора. Следовательно, к результатам классификации, получаемым при $\varphi = l_\alpha \circ \pi, \alpha > \alpha^*$, можно относиться с большим доверием, чем к тем, которые получены, например, при выборе $\varphi = l_\alpha \circ \pi, \alpha \leq \alpha^*$, когда наблюдается существенная зависимость работы алгоритма от α . Заметим, что алгоритмы, созданные специально для решения задач в нечисловых факторных пространствах (например, приводимые ниже), обладают свойством φ -инвариантности по построению, и сказанное в этом примере к ним не относится.

Важно подчеркнуть, что приведенные рассуждения ни в коей мере не доказывают возможность использования числовых методов классификации данных для решения классификационных задач в качественном факторном пространстве. Формально использовать такие методы можно, отобразив предварительно качественные показатели в числовые, но только осознавая необоснованность таких действий и возможность существенного искажения конечного результата. С другой стороны, если имеющиеся вычислительные ресурсы позволяют, можно предпринять попытку установить множество $\{\varphi\}$ биективных преобразований факторного в числовое пространство, для которых наблюдается независимость результата решения задачи от любого отдельного преобразования, взятого из этого множества. К полученному таким образом решению, в зависимости от широты определенного класса преобразований $\{\varphi\}$, можно относиться с большим или меньшим доверием. Тот факт, что результат работы алгоритма сохраняется одинаковым для широкого класса возможных преобразований $\{\varphi\}$, означает независимость числовых образов качественных показателей от порядка присвоенных им числовых значений и отражает в какой-то степени качественную природу данных.

Существуют и другие подходы к решению задач классификации в качественном пространстве, например, метод, основанный на анализе сумм мер Дайса (Семкин и др., 1986). Решающее правило метода относит классифицируемый объект к тому классу, для которого отмечается наибольшая сумма коэффициентов Дайса, вычисленных по обу-

чающей выборке для каждой градации фактора, такой же, как и у анализируемого объекта. Поскольку коэффициенты Дайса вычисляются для всевозможных пар класс/градация фактора, и заключения делаются из анализа их сумм, то очевидно, метод не требует назначения градациям факторов числовых значений и, таким образом, обладает свойством φ -инвариантности.

Ещё одним методом, используемым в сравнительном анализе решений задачи восстановления лесной растительности, является так называемый наивный байесовский классификатор (Kuncheva, 2004). Это элементарный метод, применяемый для классификации данных в качественном факторном пространстве. Рассмотрим его более подробно.

В качестве интересующей характеристики наиболее вероятного проявления фитоценоза в заданных условиях среды может выступать условная вероятность $P(A_k | B_{1i_1} B_{2i_2} \dots B_{si_s})$, где s – количество факторов среды, влияющих на формирование растительного покрова; A_k – событие, отождествляемое с конкретным проявлением фитоценоза (например, если производится восстановление по доминанту древостоя, то в качестве A_k выступает k -й доминирующий вид); B_{ij} – j -ая градация i -го фактора среды. Таким образом, $A | B_{1i_1} B_{2i_2} \dots B_{si_s}$ представляет собой событие, состоящее в том, что данное проявление лесного таксона A имеет место в условиях среды $B_{1i_1} B_{2i_2} \dots B_{si_s}$.

На практике при больших значениях параметра s определение

$P(A_k | B_{1i_1} B_{2i_2} \dots B_{si_s})$ по эмпирической выборке достаточно большого объема весьма непростая задача. Поэтому с целью упрощения процедуры ее решения рассмотрим подход, заключающийся в сведении обозначенной проблемы многомерного анализа к ряду одномерных задач.

Предположим, что

$P(A_k | B_{1i_1} B_{2i_2} \dots B_{si_s})$ можно представить в виде произведения

$P(A_k | B_{1i_1} B_{2i_2} \dots B_{si_s}) = P(A_k | B_{1i_1}) P(A_k | B_{2i_2}) \dots P(A_k | B_{si_s})$. Это предположение о взаимной независимости событий (наивное предположение, откуда и название метода) позволяет существенно упростить вычисление $P(A_k | B_{1i_1} B_{2i_2} \dots B_{si_s})$.

По величине

$P(A_k | B_{1i_1}) P(A_k | B_{2i_2}) \dots P(A_k | B_{si_s})$ можно

судить о вероятности появления A_k в условиях среды, характеризуемых совокупностью градаций $B_{1i_1} B_{2i_2} \dots B_{si_s}$. Процедуру восстановления лесной растительности в свете изложенного представим в следующем виде.

Каждой исследуемой точке (x_0) поставим в соответствие вектор $P(x_0) = \{P_j\}_{j=1}^M$ (M – максимальное число градаций явлений), j -ая компонента которого отождествляется с выражением

$P(A_j | B_{1i_1}) P(A_j | B_{2i_2}) \dots P(A_j | B_{si_s})$,

i_1, i_2, \dots, i_s – градации факторов в классифицируемой точке x_0 . Номер j^* , отвечающий максимальной компоненте вектора $P(x_0)$,

вычисленного в заданной точке x_0 , будет определять наиболее характерное для данных условий среды явление – A_{j^*} – т.е. наиболее вероятное при данных градациях i_1, i_2, \dots, i_s факторов среды B_1, B_2, \dots, B_s .

Следует отметить, что допущение о взаимной независимости условных вероятностей в наивном байесовском классификаторе на практике, как правило, не выполняется, что приводит к существенным ошибкам при его использовании (Kuncheva, 2004). Поскольку метод может быть использован для классификации данных в нечисловых факторных пространствах он, как следствие, является φ -инвариантным.

Завершая изложение теоретической части работы, перейдем к сравнительному анализу перечисленных методов на примере задачи восстановления лесной растительности учебно-опытного хозяйства "Дальневосточный" (Приморский край, Уссурийский район).

Решение задачи восстановления лесной растительности

В качестве данных обучающей выборки для решения задачи восстановления лесной растительности используем материалы лесоустройства учебно-опытного хозяйства «Дальневосточный» (далее лесхоз), выполненные по первому разряду точности под контролем специалистов высокой квалифика-

ции и оцифрованные для всей его территории при помощи регулярной сетки (с общим числом 1885 узлов) со стороной ячейки 500 м по земной поверхности. Каждое из 1885 описаний электронной базы данных включает следующие локальные характеристики фитоценоза: доминант и содоминант древостоя, доля их участия, возраст, высота, диаметр, количество древесных видов, бонитет, тип леса, полнота и запас древостоя, доминант и содоминант подроста, подлесок, численность подроста, густота подлеска и геоморфологические факторы: высота над уровнем моря, крутизна и экспозиция склона. Восстановление фитоценозов в каждой ячейке будем проводить по факторам геоморфологической группы и фитоценолотическому фактору – доминанту подроста.

Для сравнения эффективности методов оценки количества ошибочно классифицированных данных будем использовать метод скользящего экзамена (Прикладная статистика..., 1989). Согласно этому методу, имеющиеся данные (N описаний) разделяются на два множества: первое, состоящее из $N - 1$ элементов и служащее непосредственно для обучения алгоритма классификации и второе, тестовое одноэлементное множество, используемое для проверки алгоритма. Такой процесс разбиения можно проделать с каждым элементом выборки и таким образом выполнить N экспериментов тестовой классификации «бегущего» объекта, принадлежащего одноэлементному множеству. После того, как все N экспериментов выполнены, мы можем вычислить отношение правильно классифицированных случаев к общему числу экспериментов и получить представление об ошибке классификации используемого метода.

Поскольку два фактора (экспозиция склона и доминант подроста) имеют качественные градации, для возможности использования метода линейной дискриминантной классификации (*LDC*-метод) и алгоритма k -ближайших соседей сопоставим каждой такой градации натуральное число – ее порядковый номер. Чтобы оценить долю ложно классифицированных случаев, обусловленную присвоением числовых значений качественным градациям, рассмотрим (таблица) ряд экспериментов при различных преобразованиях порядковых номеров качественных градаций. Еще раз подчеркнем незаконность применения указанных алгоритмов (особенно метода линейной дискриминантной классификации) для решения подобного рода задач. Но ис-

ключительная ориентация настоящего исследования на принципиальную сторону вопроса – «что делать, если для классификации данных в качественном факторном пространстве «под рукой» имеются лишь методы, предназначенные для решения задач в числовых факторных пространствах?» – делает такие манипуляции допустимыми.

Согласно таблице, наибольшая ошибка (если судить по значениям в графе "1" – при тождественном преобразовании порядковых номеров градаций экспозиции и доминанта подроста) соответствует методу линейной дискриминантной классификации. Доля ошибочно классифицируемых этим методом объектов варьирует в зависимости от способа назначения (вида преобразования) числовых значений качественным показателям в пределах 4%, что указывает на некоторый резерв метода и на то, что подобрав определенное преобразование, можно повысить точность классификации. Существование «точностного резерва» наблюдается и для метода k -ближайших соседей. Ошибка классификации этим методом существенно зависит от вида использованного преобразования и обуславливается принятой метрикой при его реализации. Вместе с тем алгоритм k -ближайших соседей обладает высоким классификационным потенциалом (в таблице 1 он лидирует по точности верно классифицированных объектов) и тогда, когда внутриклассовые распределения существенно отличаются от нормальных (Кунчева, 2004). Последние два метода независимы от вида преобразования (φ -инвариантны) и вполне пригодны для решения классификационных задач в качественных факторных пространствах.

Высокие ошибки решения задачи для всех тестируемых методов связаны со спецификой лесотаксационных описаний, сложными перекрытиями классов в факторном пространстве и малой информативностью имеющих факторов по отношению к классифицируемым ценозам.

В качестве примера рассмотрим результаты восстановления доминантов древостоя методом k -ближайших соседей ($k = 20$) по 7 факторам среды: 3 геоморфологических фактора: высота над уровнем моря, крутизна и экспозиция склона + 3 экологические: богатство, увлажненность почв и температурный режим + доминант подростка как фитоценолотический фактор. Такая информационная база позволяет добиться вполне удовлетворительной (учитывая, что обучающая выборка –

Ошибки классификации при применении различных методов

Метод	Вид преобразования*				
	1 1/	x	x^2	x^3	$\log(x)$
LDC-метод	58.92%	56.55%	62.86%	62.92%	56.08%
Метод k -ближайших соседей ($k = 20$)	42.34%	43.46%	47.78%	50.98%	40.26%
Сумма мер Дайса	47.69 %	47.69 %	47.69 %	47.69 %	47.69 %
Наивный байесовский классификатор	56.18 %	56.18 %	56.18 %	56.18 %	56.18 %

Примечание: * – преобразование, примененное к числовым образам градаций факторов экспозиции склона и доминанта подроста.

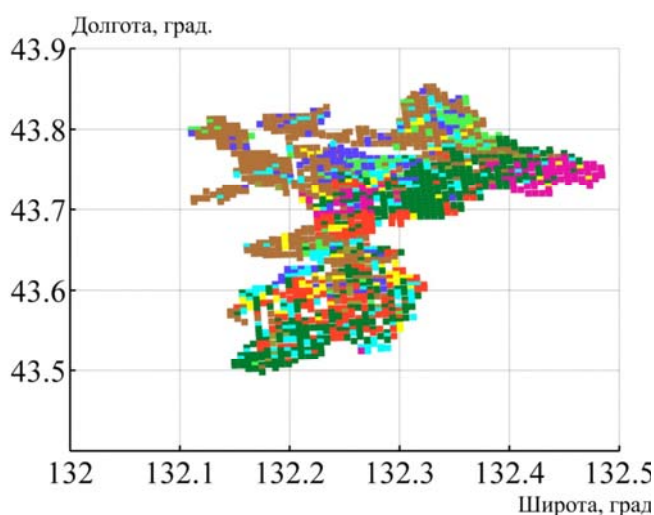


Рис. 1. Карта современного распределения лесообразующих пород (обозначения на рис. 2)

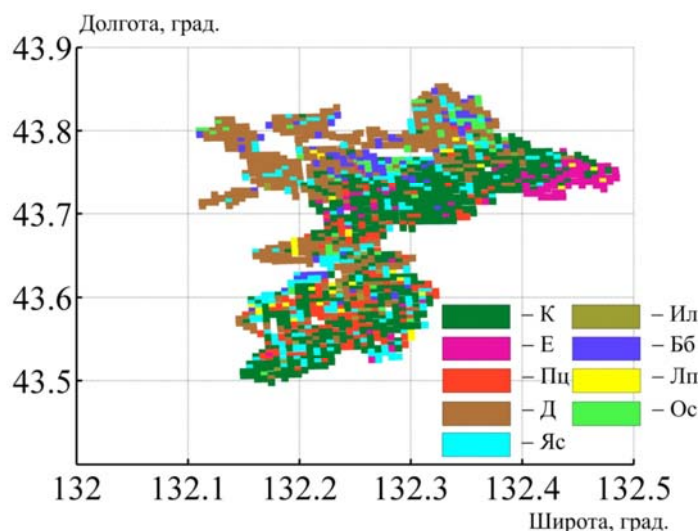


Рис. 2. Результат восстановления лесообразующих пород методом k -ближайших соседей (К – *Pinus koraiensis*^{*}, Е – *Picea jezoensis*, Пц – *Abies holophylla*, Д – *Quercus mongolica*, Яс – *Fraxinus mandshurica*, Ил – *Ulmus japonica*, Бб – *Betula platyphylla*, Лп – *Tilia amurensis*, Ос – *Populus tremula*)

Примечание: * – латинские названия указаны по сводке «Сосудистые растения советского Дальнего Востока» (1985-1996) и приводятся полностью при первом упоминании в тексте

лесотаксационные данные) точности (ошибка классификации по методу скользящего экзамена 22%) решения задачи (Брижатая, 2008). Карты исходной и восстановленной растительности представлены на рис. 1 и рис. 2 соответственно.

Из рис. 1–2 следует, что лесорастительный покров лесхоза находится в достаточно устойчивом состоянии (если не учитывать антропогенных воздействий). Тем не менее, здесь вероятны отдельные изменения, направленные в сторону увеличения ареала *Pinus koraiensis* Siebold et Zucc. в средней части лесхоза, прежде всего за счёт вытеснения им *Abies holophylla* Maxim. Также характерным является незначительное увеличение ареала *Quercus mongolica* Fisch. Ex Ledeb., преимущественно в северной и северо-западной части территории. Области, занимаемые *Picea jezoensis* (Siebold et Zucc.) Carr., остаются практически неизменными как на карте современной растительности (рис. 1), так и восстановленной (рис. 2), что свидетельствует о достаточной устойчивости данного вида и о том, что он занял свою нишу на этой территории. В северной части района исследований наблюдается незначительное замещение *Betula platyphylla* Sukacz. видом *Quercus mongolica*, сопровождающееся общим уменьшением доли вытесняемого вида на всей территории лесхоза. Необходимо осторожно относиться к прогнозным построениям в отношении этих двух видов, распространение которых во многом связано с деятельностью человека. С другой стороны, этот факт может свидетельствовать о том, что на этой части территории и ранее существовали леса с доминированием *Quercus mongolica*, то есть эти леса коренные. Кроме того, имеет место увеличение числа выделов с доминированием *Fraxinus mandshurica* Rupr. (за счёт замещения *Betula platyphylla*), что характерно для пойм и долин.

Заключение

Подводя итоги проделанной работе, отметим, что сравнительный анализ методов классификации на примере задачи восстановления лесной растительности показал эффективность метода *k*-ближайших соседей. В рассматриваемой интерпретации этот метод, строго говоря, не может применяться для

классификации данных в факторных пространствах с «нечисловыми осями», так же как и рассмотренный в работе метод линейного дискриминантного анализа. Тем не менее, назначение числовых значений грациям качественных факторов позволяет воспользоваться этими методами. При этом метод *k*-ближайших соседей характеризуется относительно высокой точностью решения задачи, несмотря на неоднозначность назначения числовых значений качественным показателям и проблему метризации факторного пространства. Специально предназначенные для решения задач классификации в нечисловых факторных пространствах методы – наивный байесовский классификатор и метод, основанный на исследовании сумм коэффициентов Дайса, – значительно уступают по точности методу *k*-ближайших соседей.

Для оценки возмущений, обусловленных назначением качественным показателям числовых значений, предлагается рассматривать результаты ряда экспериментов при различных способах назначения числовых значений качественным грациям. Если зависимость от способа выбора числовых значений существенна, необходимо с большей осторожностью относиться к получаемым в итоге результатам.

В ходе исследования установлено, что лесная растительность территории учебно-опытного хозяйства "Дальневосточный" находится в достаточно устойчивом состоянии и можно предполагать увеличение экоареала *Pinus koraiensis* в средней части района исследований (за счет постепенного вытеснения им *Abies holophylla*), увеличение экоареала *Quercus mongolica* – на севере и северо-западе, что является косвенным подтверждением его коренной природы, и увеличение числа выделов с доминированием *Fraxinus mandshurica* в поймах и на равнинах.

Л и т е р а т у р а

Брижатая А.А. Экологический анализ лесной растительности бассейна реки Комаровка (юг Приморского края): Автореф. дис. ... канд. биол. наук / Биолого-почв. ин-т ДВО РАН. – Владивосток, 2008. – 19 с.

Василевич В.И. Статистические методы в геоботанике. – Л.: Наука, 1969. – 232 с.

Петропавловский Б.С. Леса Приморского края: эколого-географический анализ. – Владивосток: Дальнаука, 2004. – 317 с.

Прикладная статистика: классификация и снижение размерности / С. А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – Москва: Финансы и статистика, 1989. – 608 с.

Рао С.Р. Линейные статистические методы и их применение. – М.: Наука, 1968. – 548 с.

Сёмкин Б.И., Петропавловский Б.С., Кошкарёв А.В., Варченко Л.И., Усольцева Л.А. О методе многомерного анализа соотношения растительности с экологическими факторами // Ботан. журн., 1986. – Т. 71. – № 9. – С. 1167–1182.

Сосудистые растения советского Дальнего Востока / Под ред. С.С. Харкевича. – Л., СПб.: Наука, 1985–1996 – Т. 1–8.

Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка, М.С. Олдендерфер, Р.К. Блэшфилд. – М.: Финансы и статистика, 1989. – 215 с.

Härdle W., Hlavka Z. Multivariate Statistics: Exercises and Solutions Springer, 2007. – 368 p.

Keller W.J. Statistical via Personal Computers // CompStat-86 Proceedings in Computational Statistics, – Wien, 1986. – P. 332–337

Kuncheva L.I. Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons, 2004. – 350 p.

A.A. Brizhataya, D.E. Kislov . **Comparative analysis of methods of reconstruction forest structure by environmental factors**

Botanical Garden-Institute FEB RAS, Vladivostok

The Institute of Automation and Control Processes FEB RAS, Vladivostok

Key words: reconstruction forest structure, classification, linear discriminate analysis, k-nearest neighbor classifier.

Comparative analysis of methods used to solve a problem of reconstruction forest structure by taxonomic data is carried out. The technique allowing to estimate the error of solution caused by using numerical methods to classify objects in non-numerical factorial spaces is suggested. For territory of the Dalnevostochnii Forest District (Ussuriiskii Raion, Primorski Krai of the Russian Far East) the most probable succession changes in tree layer are obtained.

П. 2. Табл. 1. Bibl. 11.