

МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ NDVI В ЗАДАЧЕ РАСПОЗНАВАНИЯ ТИПОВ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР ПО КОСМИЧЕСКИМ СНИМКАМ

Н.С. Воробьева, А.В. Чернов

Самарский государственный аэрокосмический университет им. академика С.П. Королёва (национально исследовательский университет), Самара, Россия

В работе рассматриваются вопросы моделирования временных рядов NDVI и применения моделированных данных в задаче распознавания типов сельскохозяйственных культур по космическим снимкам. Моделирование производится для шести типов (классов) сельскохозяйственных культур внутри каждой сельскохозяйственной зоны, расположенной на территории Самарской области. Параметры моделирования каждого класса рассчитываются по коэффициентам аппроксимации временных рядов реальных сельскохозяйственных полей функцией определенного вида. Сгенерированные наборы моделированных временных рядов используются в задаче распознавания реальных полей, расположенных на территории Самарской области.

Ключевые слова: временной ряд, вегетационный индекс, NDVI, космические снимки, распознавание сельскохозяйственных культур, алгоритм вычисления оценок, аппроксимация временных рядов, моделирование временных рядов.

Введение

Распознавание сельскохозяйственных культур по космическим снимкам в настоящее время является важной задачей применения данных дистанционного зондирования в сфере сельского хозяйства. Разнообразие методов решения данной задачи велико. Останемся на методах, использующих временные ряды вегетационных индексов, построенные по серии космических снимков. В них чаще всего используют снимки низкого пространственного разрешения, но высокого временного разрешения (съемка территории минимум один раз в сутки), что позволяет их применять для оперативного мониторинга территорий большой площади. Именно такие методы и используются в региональных геоинформационных системах сферы АПК (далее – ГИС АПК) [1] для контроля сельскохозяйственных угодий.

Качество любого метода распознавания зависит от вероятностей верной классификации объектов классов. Очевидно, для метода распознавания культур с использованием временных рядов указанный критерий качества будет напрямую зависеть от таких факторов: даты проведения классификации (чем длиннее временные ряды, тем вероятность выше); размера обучающей выборки и отсутствия в ней ошибок.

Типичной для региональных ГИС АПК является ситуация, когда фермеры декларируют данные о типах посевов на полях, однако в этой обучающей выборке присутствуют существенные ошибки и приписки. Информация может поступить слишком поздно – близко к концу или после окончания посевного сезона. Такую выборку невозможно использовать как базу для оценки качества классификации, а только для получения общих оценок распределения в пространстве признаков. В данной работе предлагается метод моделирования временных рядов и исследуется возможность применения сгенерированных

по параметрам моделирования временных рядов в качестве обучающей выборки в методе распознавания сельскохозяйственных культур.

В работе используются реальные данные о границах полей, расположенных в трех сельскохозяйственных зонах: северной, центральной и южной на территории Самарской области, и посеянных на них культурах за 2011-2015 годы. Процедуры распознавания культур, моделирования и генерации временных рядов выполняются для каждой сельскохозяйственной зоны (далее - зоны), однородной по климатическим условиям.

В данной работе рассматривается разбиение культур на следующие классы: озимые (сокращенно – оз), ранние яровые (ря), поздние яровые (пя), пар, многолетние травы (мт), неиспользуемые земли (нз).

Будем рассматривать временные ряды наиболее популярного вегетационного индекса NDVI [2].

1. Моделирование временных рядов

Можно рассматривать временной ряд как смесь полезного (идеального) сигнала и аддитивного шума, имеющую пропуски в ежедневных наблюдениях по причине облачности. Для того, чтобы использовать временные ряды как обучающую выборку, будем моделировать только полезный сигнал, так как шум и пропуски приводят к ухудшению качества классификации. Возникает вопрос о выборе вида полезного сигнала.

Широко используемые методы реконструкции временных рядов предлагают аппроксимировать временные ряды функциями такого вида: асимметричная гауссова [3,4], двойная логистическая [5]. Было проведено сравнение качества моделирования указанных вариантов и многочленов различного вида по качеству аппроксимации, гибкости и устойчивости модели. Подробное описание результатов сравнения выходит за рамки настоящей работы. В итоге для моделирования полезной составляющей временного ряда (назовем ее «идеальной кривой») некоторого класса предлагается использовать следующую функцию:

$$F(x) = (x-a)(x-b)P_n(x) + kx + e, \quad (1)$$

где $P_n(x)$ – линейная комбинация полиномов Лежандра до n -ой степени включительно, a – самая ранняя дата, когда хотя бы у одного объекта класса в данной зоне есть наблюдение, b – самая поздняя дата, когда хотя бы у одного объекта класса в данной зоне еще есть наблюдения. Коэффициенты k, e получают из условий:

$$\begin{aligned} k \cdot a + e &= y_a \\ k \cdot b + e &= y_b, \end{aligned} \quad (2)$$

где y_a, y_b – значения усредненного тренда класса, аппроксимированного полиномом вида $P_n(x)$, в точках a, b соответственно. Усредненный тренд класса рассчитывается с использованием всех временных рядов класса и состоит из усредненных значений для каждого

дня наблюдения. Таким образом, для некоторого класса параметры k, e будут общими для всех кривых данного класса, а различаться будут коэффициенты полинома $P_n(x)$, иначе говоря, каждая идеальная кривая характеризуется следующим вектором коэффициентов: $\bar{p} = (p_1, p_2, \dots, p_{n+1})^T$.

Далее будем рассматривать $n=4$. Определившись с видом кривой, переходим к следующему этапу – расчет параметров моделирования идеальных кривых.

2. Расчет параметров моделирования

Для расчета параметров будем использовать временные ряды, построенные по данным со спутника Terra/MODIS, обработанным до уровня продукта MOD09GQ, соответствующие реальным полям. Общее количество реальных временных рядов – 20940. Для данных рядов имеется соответствие «временной ряд» – «класс культур». Далее в таблице 1 приведено общее количество реальных временных рядов для каждого класса:

Табл. 1. Общее количество реальных временных рядов для разных классов

Класс	Количество реальных временных рядов
многолетние травы	887
неиспользуемые земли	2416
озимые	2972
пар	4650
поздние яровые	5969
ранние яровые	4026

С учетом того, что профили временных рядов значительно меняются в зависимости от года, зоны и культуры, процедура оценки параметров производится отдельно для каждой культуры в выбранной зоне в заданном году. Для зафиксированной тройки «культура-зона-год» расчет параметров моделирования состоит из следующих этапов.

1. Аппроксимация временных рядов функцией вида (1) по методу наименьших квадратов. В результате получаем некоторое количество реализаций (по количеству временных рядов) вектора коэффициентов \bar{p} , имеющего многомерный нормальный закон распределения.
2. Расчет вектора матожиданий \bar{M} и ковариационной матрицы B , характеризующих закон распределения вектора \bar{p} .

Для Самарской области за 2011-2015 годы рассчитано 90 различных параметров моделирования идеальных кривых для заданной формулой (1) функции.

Стоит отметить, что получающиеся на первом этапе невязки аппроксимации – это шум, вычтенный из полезного идеального сигнала. При аппроксимации функцией вида (1) для каждой тройки «культура-зона-год» получают некоррелированный шум. Такой вывод можно сделать, проанализировав значения АКФ, построенной по последовательности невязок для выбранной тройки «культура-зона-год». Все значения АКФ кроме соответ-

ствующего нулевому смещению не превышают значения 0,5, то есть можно говорить о некоррелированности значений. Это значит, что выбрана достаточная степень полинома в функции (1), и значения невязки не содержат в себе неучтенных функцией (1) остатков полезного сигнала.

Среднее значение среднеквадратической ошибки аппроксимации временных рядов для тройки «культура-зона-год» составляет 0,047. На рисунке приведен пример аппроксимации временного ряда функцией вида (1).

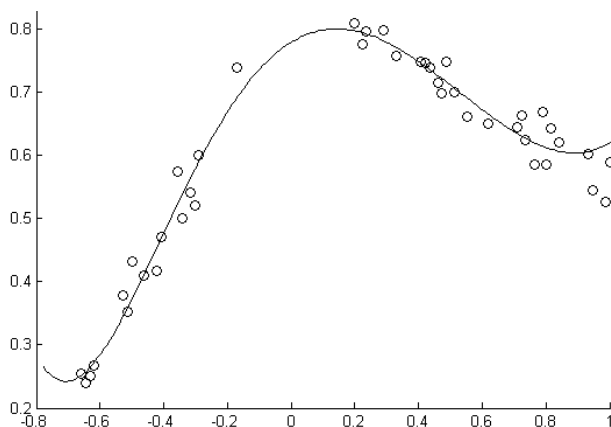


Рис. 1. Пример аппроксимации временного ряда

На рисунке по оси X отложена временная координата, дата наблюдения временного ряда, которая для удобства расчетов была переведена в диапазон [-1;1]. По оси Y отложены значения индекса NDVI.

3. Генерация «идеальных кривых»

«Идеальная кривая» определяется вектором коэффициентов \bar{p} , поэтому для ее получения необходимо сгенерировать вектор \bar{p} , имеющий многомерный нормальный закон распределения с заданными вектором матожиданий \bar{M} и ковариационной матрицей B , следующим образом [6]:

1. Сгенерировать вектор $\bar{\xi} = (\xi_1, \dots, \xi_{n+1})^T$, компоненты которого являются независимыми случайными величинами, имеющими стандартный нормальный закон распределения.
2. Вычислить матрицу A , представляющую собой разложение Холецкого ковариационной матрицы B .
3. Рассчитать вектор коэффициентов через линейное преобразование вектора $\bar{\xi}$:

$$\bar{p} = A\bar{\xi} + \bar{M}$$

Далее рассчитываются ежедневные значения идеальной кривой по полученному вектору коэффициентов \bar{p} и набору параметров k, e , общих для тройки «культура-зона-год» в допустимом диапазоне $[a, b]$.

По указанному алгоритму были сгенерированы идеальные кривые для шести типов культур для трех зон Самарской области за 2011-2015 годы. Для каждой тройки «культура-зона-год» было сгенерировано 4000 кривых.

4. Алгоритм распознавания типов культур

Для распознавания типов культур с использованием временных рядов будет использован метод на основе алгоритма вычисления оценок (АВО) [7]. Преимуществом АВО является возможность классификации объектов, в признаках которых имеются пропуски. Как было сказано выше – пропуски в значениях временных рядов, а значит и в наборе признаков, возникают по причине облачности. Конкретизируем модель АВО для распознавания типов культур:

1. Признаки – значения временного ряда.
2. Система опорных множеств признаков состоит из единственного множества, включающего в себя все признаки.
3. Функция близости распознаваемого a и эталонного ω объектов рассчитывается по формуле:

$$\rho = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (\omega_n - a_n)^2}, \quad (3)$$

где $a_n, n = \overline{0, N-1}$ – признаки объекта a (набор значений временного ряда); $\omega_n, n = \overline{0, N-1}$ – признаки объекта ω . Функция близости вычисляется только по тем дням n , в которые и у объекта ω и у объекта a имеются значения временных рядов ω_n и a_n , соответственно.

4. Значение функции близости $f(\omega, a)$ между эталонным объектом ω и распознаваемым

объектом a вычисляется так: $f(\omega, a) = \begin{cases} 1, & \text{если } \rho \leq T \\ 0, & \text{если } \rho > T \end{cases}$, где T – пороговое значение близости.

5. Оценка $\Gamma(\Omega_j)$ близости объекта a к некоторому классу Ω_j вычисляется так:

$$\Gamma(\Omega_j) = \sum_{\omega \in \Omega_j} f(\omega, a)$$

6. Классификация распознаваемого объекта a будет произведена в класс Ω_c в соответствии с решающим правилом

$c = \arg \max_{m=0, M-1} \Gamma(\Omega_m)$, M – количество классов.

5. Экспериментальные исследования

Далее приводятся результаты экспериментов, проведенных для анализа целесообразности использования моделированных временных рядов в качестве обучающих данных в методах распознавания вместо временных рядов, построенных для реальных полей. Первые три эксперимента использовали моделированные временные ряды в качестве эта-

лонной выборки в АВО и были проведены для распознавания реальных полей, расположенных в Самарской области в 2014 году в северной, центральной и южной зоне, соответственно. В четвертом эксперименте в качестве обучающих данных использовались временные ряды, построенные для реальных полей, а распознавание проводилось без разделения полей по зонам. Это связано с небольшим количеством реальных данных, которых недостаточно в случае классификации по зонам для формирования и обучающей выборки, и набора распознаваемых полей. Для каждого эксперимента приведена матрица вероятностей классификации из класса в класс в таблицах 2-5 соответственно.

1. Северная зона (общее количество распознаваемых объектов $K=1341$). Общая вероятность верной классификации $Q=0,58$.

Табл. 2. Матрица вероятностей классификации из класса в класс. 2014 год. Зона «север»

Класс	мт	нз	оз	пар	ря	пя
мт	0,33	0,51	0,05	0,01	0,05	0,04
нз	0,13	0,73	0,02	0,07	0,05	0,01
оз	0,03	0,10	0,74	0,08	0,04	0,01
пар	0,07	0,30	0,04	0,44	0,09	0,05
ря	0,02	0,13	0,04	0,07	0,56	0,16
пя	0,03	0,11	0,02	0,04	0,14	0,66

2. Центральная зона ($K=2324$). $Q=0,73$

Табл. 3. Матрица вероятностей классификации из класса в класс. 2014 год. Зона «центр»

Класс	мт	нз	оз	пар	ря	пя
мт	0,25	0,53	0,11	0,08	0,01	0,02
нз	0,03	0,88	0,02	0,04	0,01	0,02
оз	0,03	0,05	0,87	0,03	0,01	0,01
пар	0,03	0,25	0,05	0,55	0,06	0,05
ря	0,01	0,14	0,01	0,06	0,68	0,11
пя	0,00	0,12	0,01	0,03	0,07	0,76

3. Южная зона ($K=2766$). $Q=0,70$.

Табл. 4. Матрица вероятностей классификации из класса в класс. 2014 год. Зона «юг»

Класс	мт	нз	оз	пар	ря	пя
мт	0,44	0,39	0,06	0,03	0,06	0,03
нз	0,12	0,83	0,01	0,02	0,00	0,01
оз	0,06	0,02	0,87	0,02	0,01	0,02
пар	0,07	0,14	0,03	0,56	0,09	0,10
ря	0,03	0,05	0,01	0,08	0,68	0,14
пя	0,01	0,03	0,02	0,09	0,14	0,70

4. Классификация реальных данных по реальным данным. Оценка качества классификации проводилась с помощью скользящего контроля. Выборка из всех реальных временных рядов за 2014 год пять раз разбивалась на обучающую (эталонные объекты) и контрольную (распознаваемые объекты) в соотношении 2:1. Среднее значение $Q=0,69$.

Табл. 5. Матрица вероятностей классификации из класса в класс. 2014 год

Класс	мт	нз	оз	пар	ря	пя
мт	0,40	0,42	0,08	0,03	0,01	0,03
нз	0,17	0,77	0,01	0,02	0,02	0,01
оз	0,07	0,03	0,86	0,01	0,01	0,02
пар	0,12	0,19	0,03	0,53	0,06	0,07
ря	0,04	0,09	0,01	0,04	0,68	0,14
пя	0,03	0,07	0,01	0,07	0,07	0,75

Таким образом, приведенные выше результаты экспериментов позволяют сделать вывод о возможности использования моделированных временных рядов в качестве обучающей выборки.

Заключение

В данной работе продемонстрирована идея использования моделированных временных рядов в качестве обучающей выборки в методе распознавания культур. Эта идея имеет следующие достоинства:

1. Генерация обучающей выборки любого объема по параметрам моделирования.
2. Возможность использования набора моделированных временных рядов как базу для оценки качества классификации.
3. Возможность улучшения качества классификации за счет подбора различных функций аппроксимации.
4. Накопив некоторую статистику реальных данных за ряд лет, можно сгенерировать по ним модельные временные ряды и далее использовать их как некоторый исторический шаблон развития культур. В начале сезона, когда обучающая выборка мала или отсутствует, но есть некоторое количество полей с достоверной информацией о посеянных культурах (назовем их опорные поля), из набора исторических шаблонов можно выбрать шаблон, наиболее близкий по развитию культур к опорным полям. Далее сгенерированные временные ряды выбранного исторического шаблона можно использовать как обучающую выборку в методе распознавания типов культур в текущем году.

Благодарности

Работа выполнена при финансовой поддержке РФФИ, грант № 16-37-00043_мол_а «Разработка методов использования данных геоинформационных систем для обработки данных дистанционного зондирования Земли».

Литература

1. Vorobiova, N.S. How to Use Geoinformation Technologies and Space Monitoring for Controlling the Agricultural Sector in Samara Region / N.S. Vorobiova, A.Yu. Denisova, A.V. Kuznetsov, A.M. Belov, A.V. Chernov, V.V. Myasnikov // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. – 2015. – Vol. 25(2). – P. 347–353.

2. Майорова, В.И. Контроль состояния сельскохозяйственных полей на основе прогнозирования динамики индекса NDVI по данным космической мультиспектральной и гиперспектральной съёмки / В.И. Майорова, А.М. Банников, Д.А. Гришко, И.С. Жаренов, В.В. Леонов, А.Г. Топорков, А.А. Харлан // Наука и образование: научное издание МГТУ им. Н.Э. Баумана. – 2013. – В. 07. – С. 199-228. – ISSN 1994-0408.
3. Ozdogan, M., The spatial distribution of crop types from MODIS data: Temporal unmixing using Independent Component Analysis / Mutlu Ozdogan // Remote Sensing of Environment. – 2010. – Vol. 114(6). – P. 1190–1204.
4. Fischer, A. A Model for the Seasonal Variations of Vegetation Indices in Coarse Resolution Data and Its Inversion to Extract Crop Parameters / Alberte Fischer // Remote Sensing of Environment. – 1994. – Vol. 48. – P. 220-230.
5. Wei, W. Selecting the optimal NDVI time-series reconstruction technique for crop phenology detection / Wei Wei, Wenbin Wu, Zhengguo Li, Peng Yang, Qingbo Zhou // Intelligent Automation & Soft Computing. – DOI:10.1080/10798587.2015.1095482.
6. Моделирование экспериментальных данных для решения задач распознавания образов. Методические указания к лабораторной работе №1 / Э.И. Коломиец, В.В. Мясников. – Самара, 2010. – 20 с.
7. Воробьева, Н.С. Распознавание сельскохозяйственных культур по космическим снимкам с использованием алгоритма вычисления оценок / Н.С. Воробьева // Материалы Международной конференции и молодежной школы «Информационные технологии и нанотехнологии (ИТНТ-2015)». – 2015. – С.83-88.