

РАСПОЗНАВАНИЕ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР ПО КОСМИЧЕСКИМ СНИМКАМ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ВЫЧИСЛЕНИЯ ОЦЕНОК

Н.С. Воробьева

Самарский государственный аэрокосмический университет им. академика С.П. Королёва

В статье предлагается метод распознавания сельскохозяйственных культур по космическим снимкам, основанный на алгоритме вычисления оценок. Признаками для классификации являются набор значений временного ряда, построенного по последовательности космических снимков и географическая координата поля – широта. Тестирование метода выполнялось по снимкам Terra/MODIS с использованием наземной информации. Приводится сравнение качества классификации предлагаемого метода с классификатором по расстоянию Махаланобиса.

1. Введение

В настоящее время управление в сфере АПК является областью активного развития и внедрения методов, применяющих данные дистанционного зондирования (далее – ДДЗ) для решения производственных задач [1]. Одной из таких задач является распознавание сельскохозяйственных (далее – с/х) культур по космическим снимкам. Ее решение крайне важно для оперативного мониторинга и контроля использования пахотных земель, проверки сведений, предоставляемых сельхозпроизводителями о засевах полей культурами, а также для составления карты использования земель под культурами, в зонах, где информация от сельхозпроизводителей отсутствует.

Предлагается большое число методов для решения задачи распознавания с/х культур по космическим снимкам [2-6]. Наибольшую важность представляют методы, подходящие для применения на больших территориях (регионального масштаба и более). Территории большой площади, в отличие от территорий хозяйств или групп хозяйств, затруднительно обследовать наземным способом, а ДДЗ предоставляют такую возможность [9, 10].

В данной работе предлагается метод распознавания с/х культур, основанный на алгоритме вычисления оценок и подходящий к использованию на территориях регионального масштаба. Метод позволяет учитывать географическое положение поля и использовать временные ряды, в которых имеются пропуски данных. Предлагаемый метод распознавания основан на оценке близости исследуемого и эталонного объектов. Пропуски возникают из-за наличия облаков на космических снимках в некоторый день наблюдения. Возможность использования данных с пропусками не потребует дополнительно проведения процедуры интерполяции временных рядов. Качество предлагаемого метода сравнивается с качеством классификатора, основанным на расстоянии Махаланобиса.

2. Исходные данные и их предварительная обработка

Все данные для исследования получены для территории Самарской области. В работе используются космические снимки и наземные данные о полях за 2014 год. Далее приведено описание самих данных и методов их предварительной обработки.

2.1. Космические снимки

Для построения временных рядов используются данные с космического аппарата Terra, регистрируемые спектрорадиометром MODIS. В результате наземной обработки зарегистрированных на спутнике сырых данных получают так называемые продукты [7]. Продукт – есть результат некоторой обработки исходных (или уже обработанных до некоторого уровня) данных. В данной работе для построения временных рядов используются красный и ближний ИК каналы ежедневного продукта MOD09GQ. Продукт прошел радиометрическую, атмосферную коррекцию, имеет географическую привязку и

представляет собой минимальный уровень комбинирования ежедневных данных разрешения 250 м.

Предварительная обработка продуктов MOD09GQ (для каждого дня) включает в себя следующие этапы:

1. Составление мозаики 2-х тайлов (h20v03, h21v03), покрывающих территорию Самарской области. Тайл – ячейка синусоидальной сетки MODIS..
2. Попиксельная синхронизация со снимками за предыдущие дни: приведение к единой проекции и разрешению; установка одинакового размера по вертикали и горизонтали; задание единой привязки угловых точек.
3. Формирование маски облачности.

2.2. Наземные данные

Наземные данные о полях были предоставлены с/х производителями и необходимы для обучения классификатора и оценки его качества. Набор сведений для каждого поля следующий: граница и площадь поля, посеянные культуры.

Границы полей были наложены на покрытие, составленное по космическим снимкам среднего разрешения 20-30 м, с целью проверки однородности использования полей. Неоднородность обуславливается следующими факторами: присутствие нескольких культур или видов использования на поле, использование только части поля под засев культуры, неоднородность всходов. Все обнаруженные неоднородные поля были разделены на соответствующее количество однородных областей. Исследование однородности проводилось с использованием метода сегментации [9]. В результате были получены новые границы – границы участков, которые и использовались для расчета временных рядов и распознавания культур.

В данной работе все культуры и типы использования полей были разделены на следующие группы: многолетние травы, неиспользуемые земли, озимые, пар, ранние яровые, поздние яровые. Такое разбиение на классы и будет использоваться далее в алгоритме классификации на основе вычисления оценок.

2.3. Временные ряды

Для построения временных рядов был выбран нормализованный разностный вегетационный индекс NDVI. Значение индекса вычисляется по формуле:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (1)$$

где ρ_{NIR}, ρ_{RED} – значения отраженного излучения в ближней инфракрасной и красной зоне спектра, соответственно.

Для построения временных рядов индекса NDVI использовались космические снимки Terra/MODIS, полученные в период с 1 марта по 30 сентября 2014 года. Значение индекса для каждого участка вычислялось следующим образом:

- расчет усредненного по всем пикселям участка значений красного и ближнего инфракрасного значений каналов ρ_{NIR}, ρ_{RED} ;
- вычисление NDVI участка по значениям ρ_{NIR}, ρ_{RED} по формуле (1).

Значения NDVI, упорядоченные по дате получения снимка, формируют временной ряд индекса NDVI на участке.

3. Метод распознавания с/х культур

Для распознавания культур предлагается использовать метод классификации, основанный на алгоритме вычисления оценок. Класс алгоритмов распознавания, основанных на вычислении оценок (далее - АВО), предложенный Журавлевым Ю.И. [8] – это не заранее заданный алгоритм, а модель распознающего алгоритма. И в каждом конкретном случае эту модель надо конкретизировать. Далее приведено описание

предлагаемого в статье метода идентификации культур, представляющего собой конкретизацию модели АВО.

3.1. Модель АВО для идентификации культур

Исходными данными для работы алгоритма являются набор эталонных и распознаваемых объектов. Все объекты характеризуются следующими признаками: значениями временного ряда и географической координатой - широтой. Необходимо произвести классификацию набора распознаваемых объектов в заранее заданные классы.

Априорная информация задается в виде таблицы (см. таблицу 1).

Таблица 1. Априорная информация

класс	объекты	Признаки					
		x_0	x_1	x_2	...	x_{N-1}	δ
Ω_0	ω_0	$\omega_{0,0}$	$\omega_{0,1}$	$\omega_{0,2}$		$\omega_{0,N-1}$	δ_0
	ω_1	$\omega_{1,0}$	-	$\omega_{1,2}$		$\omega_{1,N-1}$	δ_1
					
Ω_1	ω_p	-	$\omega_{p,1}$	$\omega_{p,2}$		-	δ_p
					
...							

В таблице 1 используются следующие условные обозначения: x_i, δ – признаки: значение временного ряда в день i и широта объекта, соответственно; Ω_m – класс $m, m = \overline{0, M-1}$, где M – количество классов; ω_p – эталонный объект из некоторого класса Ω_m . Объект ω_p имеет следующие признаки: широту δ_p и набор значений временного ряда $\omega_{p,n}, n = \overline{0, N-1}$. Здесь $\omega_{p,n}$ – значение временного ряда объекта ω_p в день n , N – общее количество дней наблюдения. И эталонные, и распознаваемые объекты могут иметь пропуски в значениях временных рядов. Конкретизация модели АВО подразумевает задание следующих подпунктов.

Система опорных множеств признаков

Система опорных множеств признаков в данном случае состоит из единственного множества, включающего в себя все признаки.

Функция близости

Будем рассчитывать близость ρ распознаваемого объекта a и эталонного объекта ω_p следующим образом:

$$\rho = e^{-(k\rho_1 + (1-k)\rho_2)}, \quad (2)$$

где ρ_1 – величина, характеризующая близость двух объектов по значениям их временных рядов, ρ_2 – величина, характеризующая близость двух объектов по широте, k – параметр, определяющий веса ρ_1 и ρ_2 в итоговом значении близости. ρ_1 и ρ_2 вычисляются по формулам (3) и (4), соответственно.

$$\rho_1 = \sum_{n=0}^{N-1} (\omega_{p,n} - a_n)^2 \quad (3)$$

$$\rho_2 = |\delta_p - \delta_a|, \quad (4)$$

где $a_n, n = 0, N-1$ и δ_a – признаки объекта a : набор значений временного ряда и широта, соответственно. Значение ρ_1 вычисляется только по тем дням n , в которые и у объекта ω_p и у объекта a имеются значения временных рядов $\omega_{p,n}$ и a_n , соответственно.

Значение функции близости $f(\omega_p, a)$ между эталонным объектом ω_p и распознаваемым объектом a вычисляется следующим образом:

$$f(\omega_p, a) = \begin{cases} 1, & \text{если } \rho > T \\ 0, & \text{если } \rho \leq T \end{cases}, \quad (5)$$

где T – пороговое значение близости.

Оценка близости для класса

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

$$\Gamma(\Omega_j) = \sum_{\omega_p \in \Omega_j} f(\omega_p, a). \quad (6)$$

Решающее правило

Классификация распознаваемого объекта a будет произведена в класс Ω_c в соответствии с решающим правилом. Зададим два варианта решающих правил:

1) $c = \arg \max_{m=0, M-1} \Gamma(\Omega_m)$; 2) $c = \arg \max_{m=0, M-1} \frac{\Gamma(\Omega_m)}{q_m}$, q_m – количество элементов в классе Ω_m .

4. Результаты

4.1. Набор объектов для тестирования

Для оценки качества предлагаемого классификатора была сформирована тестовая выборка из 6424 участков. Оценка качества проводилась с помощью скользящего контроля. Тестовая выборка пять раз разбивалась на обучающую и контрольную выборки в соотношении 2:1.

4.2. Классификатор с использованием АВО

Для каждого варианта решающего правила выполнялся поиск значений параметров k, T , на которых достигается наилучший результат классификации. Поиск производился методом перебора с шагом 0.01 для k , и 0.001 для T . Для решающего правила №1 наилучшие результаты классификации достигаются при $k = 0.98$, $T = 0.9945$. Для решающего правила №2 наилучшие результаты достигаются при $k = 0.97$, $T = 0.9945$. Результаты классификации из класса с использованием АВО приведены в таблице 2.

Таблица 2. Результаты классификации из класса в класс для алгоритма на основе АВО

Название класса	Решающее правило №1, $k = 0.98$, $T = 0.9945$, $Q = 0.72$						Название класса	Решающее правило №2, $k = 0.97$, $T = 0.9945$, $Q = 0.69$					
	мт	нз	оз	пар	ря	пя		мт	нз	оз	пар	ря	пя
мн-е травы (мт)	0.18	0.54	0.13	0.06	0.02	0.03	мт	0.40	0.42	0.08	0.03	0.01	0.03
неисп. зем. (нз)	0.02	0.87	0.02	0.03	0.04	0.01	нз	0.17	0.77	0.01	0.02	0.02	0.01
озимые (оз)	0.00	0.04	0.89	0.02	0.02	0.02	оз	0.07	0.03	0.86	0.01	0.01	0.02
пар	0.01	0.19	0.04	0.58	0.08	0.08	пар	0.12	0.19	0.03	0.53	0.06	0.07
ран. яров. (ря)	0.00	0.08	0.01	0.04	0.69	0.17	ря	0.04	0.09	0.01	0.04	0.68	0.14
поздн. яров. (пя)	0.00	0.07	0.01	0.07	0.06	0.78	пя	0.03	0.07	0.01	0.07	0.07	0.75

На рисунке 1 приведены графики зависимости общей вероятности верной классификации Q от порога T . При $k = 1$ широта не учитывается. Из графиков видно, что значения Q выше для варианта, когда широта учитывается.

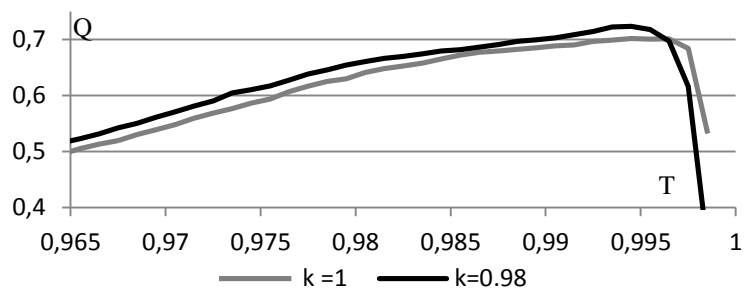


Рисунок 1 – Зависимость вероятности верной классификации Q от порога T для k=1 и k=0.98

4.3. Классификатор по расстоянию Махаланобиса

Для оценки качества предлагаемого в статье классификатора на основе АВО была произведена классификация тестовой выборки по расстоянию Махаланобиса [9] с использованием указанной выше схемы скользящего контроля. Результаты классификации из класса в класс – в таблице 3. ОВВК – 0.64.

Таблица 3. Результаты классификации из класса в класс для классификатора по расстоянию Махаланобиса

	мт	нз	оз	пар	ря	пя
мн-е травы (мт)	0.51	0.22	0.11	0.09	0.06	0.01
неисп. зем. (нз)	0.10	0.70	0.02	0.07	0.07	0.04
озимые (оз)	0.08	0.05	0.75	0.05	0.04	0.04
пар	0.06	0.11	0.04	0.61	0.09	0.09
ран. яров. (ря)	0.06	0.06	0.03	0.11	0.55	0.19
поздн. яров. (пя)	0.03	0.04	0.02	0.10	0.12	0.68

5. Заключение

В статье предлагается метод распознавания сельскохозяйственных культур на основе алгоритма вычисления оценок. Преимущества предложенного алгоритма: использование временных рядов с пропусками и учет географического положения поля. Как видно из представленных результатов классификации, метод АВО дает общую вероятность верной классификации ($Q=0.72$) выше, чем классификатор по расстоянию Махаланобиса (0.64). Однако, значение 0.72 не является удовлетворительным, и требуется улучшение качества классификации. Детальное изучение тестовой выборки показало, что в данных имеются недостоверности – указана неверная культура для поля, а также деление культур на группы было выполнено неоптимальным образом. Поэтому направлением дальнейших исследований будет изучение вопроса деления культур на группы, так чтобы наилучшее качество классификации достигалось на максимально возможном разбиении культур на классы.

Литература

1. Отраслевые решения для сельского хозяйства/ <http://sovzond.ru/industry-solutions/agriculture/>
2. Wardlow B.D., Egbert S.L.: Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains//Remote sensing of environment, 2008, Vol. 112, No 3, pp. 1096-1116.
3. Wardlow B.D.: An Evaluation of Time-Series MODIS 250-meter Vegetation Index Data for Crop Mapping in the U.S. Central Great Plains – Michigan: ProQuest Information and Learning Company, 2005.
4. Brown J.C., Kastens J.H., Coutinho A.C., Victoria D.C., Bishop C. R.: Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data// Remote Sensing of Environment, 2013, Vol. 130, pp. 39–50.
5. Плотников Д.Е., Барталев С.А., Жарко В.О., Михайлов В.В., Просяникова О.И.: Экспериментальная оценка распознаваемости агрокультур по данным сезонных спутниковых измерений спектральной яркости // Современные проблемы дистанционного зондирования земли из космоса, 2011, Т.8, №1, С199-208.
6. Плотников Д.Е., Барталев С.А., Лупян Е.А.: Признаки распознавания пахотных земель на основе многолетних рядов данных спутникового спектрометра MODIS. Современные проблемы дистанционного зондирования Земли из космоса, 2010, Т.7, №1, С330-341.
7. MODIS Website/ <http://modis.gsfc.nasa.gov/>
8. Журавлев Ю.И., Камилев М.М., Туляганов Ш.Е.: Алгоритмы вычисления оценок и их применение. Издательство «Фан» УзССР, 1974.

9. Vorobiova N. S., Denisova A. Yu., Kuznetsov A. V., Belov A. M., Chernov A. V., Myasnikov V. V. How to Use Geoinformation Technologies and Space Monitoring for Controlling the Agricultural Sector in Samara Region // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications, 2015, Vol. 25, No. 2, pp. 347–353.

10. Воробьева Н.С., Тимбай Е.И. Разработка геоинформационной системы учета и контроля земель сельхозназначения // Компьютерная оптика, Самара-Москва, Российская Академия наук, Институт систем обработки изображений РАН, 2009, том 33, №3, С340-344.