



<http://dx.doi.org/10.15407/dopovidi2016.08.026>

УДК 528.854

Н. Н. Куссуль, А. Ю. Шелестов, Н. С. Лавренюк, И. Н. Бутко

Институт космических исследований НАН Украины и ГКА Украины, Киев

E-mail: nataliia.kussul@gmail.com, andrii.shelestov@gmail.com,

nick_93@ukr.net, butko.igor.68@gmail.com

Информационные технологии глубинного машинного обучения для анализа изменений земного покрова

(Представлено членом-корреспондентом НАН Украины О. П. Федоровым)

Предложен метод и информационная технология геопространственного анализа изменений земного покрова на основе многолетних спутниковых наблюдений. Так как решаемая проблема относится к классу задач с большим объемом данных (big data), для ее решения разработан метод глубинного машинного обучения, основанный на иерархической нейросетевой модели. Метод позволяет решать широкий класс прикладных задач анализа изменений земного покрова и землепользования.

Ключевые слова: big data, глубинное обучение, нейросетевые модели, изменение земного покрова.

Запуск программы Европейского союза Copernicus ознаменовал начало новой эры в истории мировой космонавтики. В течение 20 лет, начиная с 2014 г., в рамках этой программы будут гарантированно предоставляться бесплатные спутниковые данные высокого разрешения в операционном режиме.

Так, в течение 2015 г., получено более 500 сцен радарного спутника Sentinel-1 и несколько сотен сцен данных спутника Sentinel-2. Каждый снимок содержит миллионы пикселей данных в нескольких каналах и покрывает значительную территорию Украины. Учитывая появление в свободном доступе данных других спутников, в частности Landsat-8 и Proba-V, можно с полной уверенностью констатировать возникновение проблемы больших объемов данных (big data), для решения которой необходимы принципиально новые методы и автоматические информационные технологии.

Для решения проблемы больших объемов данных в последние годы активно развиваются методы глубинного обучения (deep learning) [1], которые, как правило, разрабатыва-

© Н. Н. Куссуль, А. Ю. Шелестов, Н. С. Лавренюк, И. Н. Бутко, 2016

ются для других предметных областей, в частности обработки текстовой и видеоинформации. Проблема обработки больших объемов спутниковых и геопространственных данных практически не изучена, хотя эта постановка задачи не является новой [2].

В основе глубинного обучения лежит методология системного анализа и иерархической обработки сложной информации. Как известно, обработка больших объемов данных и извлечение из них содержательной информации является сложной задачей, которая не может быть решена с помощью одной модели, даже очень высокой сложности. Для решения таких задач предназначены методы системного анализа [3], которые предполагают иерархическую декомпозицию сложной задачи на более простые, решаемые на основе отдельных моделей, с последующим получением синергетического результата от иерархической комбинации различных моделей. Примеры таких систем описаны в [4–6]. Поскольку проблема обработки больших объемов данных относится к области наук о данных, то ее решение строится на основе индуктивного подхода и интеллектуальных моделей, зачастую нейросетевых [7, 8]. Таким образом, глубинное обучение предполагает иерархическое извлечение осмысленной информации из больших объемов данных с помощью различных интеллектуальных моделей.

Задачу спутникового мониторинга в контексте решения проблем устойчивого развития можно рассматривать как задачу глубинного обучения, поскольку из числовых значений на матрице принимающего устройства необходимо извлечь осмысленную информацию для большого количества данных различного пространственного и временного разрешения. Нами такой подход применен к решению задачи картографирования земного покрова и оценки его изменений для территории Украины в течение 20 лет с 1990 по 2010 гг.).

Спутниковые данные. В последние годы предложено несколько вариантов карт земного покрова, в том числе высокого (30-метрового) разрешения. Большое внимание уделяется оценке их точности. Однако все существующие карты не обеспечивают достаточной точности оценок на уровне региона, так как являются глобальными и для их построения используются обучающие данные других регионов [9]. Поэтому задача создания региональных карт земного покрова и оценки его изменений является актуальной. При решении задач анализа изменений земного покрова для территории Украины за период 1990–2010 гг. на основе ретроспективных данных целесообразно использовать данные космических аппаратов семейства Landsat. Композиты каналов этого спутника используются при формировании обучающих выборок на основе фотоинтерпретации изображений. Для построения карт земного покрова за 1990, 2000 и 2010 гг. использовано 463 спутниковых изображения.

Формирование обучающей и тестовой выборок. Поскольку классификация осуществляется по ретроспективным данным, для построения обучающей выборки используется метод фотоинтерпретации. При этом в процессе построения обучающей выборки важно обеспечить пропорциональное представление всех классов земной поверхности по исследуемой территории.

В состав построенных выборок входят полигоны для шести классов, соответствующих номенклатуре LUCAS, а именно: искусственные объекты, пахотные земли (обрабатываемые земли в текущем году), лес, необрабатываемые земли, открытый грунт и вода. В результате формирования полигонов получена выборка из 13045 полигонов за 2010, 11932 — для 2000 г. и 11585 — за 1990 г. Выборка, построенная отдельно для каждой области Украины, случайным образом разбивается на два множества: обучающую (составляет 50% общего количества полигонов) и тестовую (остальные 50%).

Метод глубинного обучения. Для решения задачи оценки изменений земного покрова предлагается следующая иерархическая структура модели глубинного обучения (рис. 1). Модель включает четыре уровня: I — предобработка, II — попиксельная классификация, III — постобработка (векторизация), IV — оценка изменений земного покрова. Рассмотрим каждый из этих уровней детальнее.

I. Предобработка (фильтрация шума и кластеризация данных). При использовании оптических спутниковых данных шумом являются облака и тени, искажающие изображение. Задачу их устранения предлагается решать путем кластеризации временного ряда снимков с помощью самоорганизующейся сети Кохонена. Этот метод детально описан в [10]. На рис. 1, I (см. вклейку) красным цветом показаны входные спутниковые данные в проекции на плоскость, а синим цветом — результат кластеризации. Справа приведен пример применения этого алгоритма к оптическим изображениям спутника Landsat-8.

II. Классификация элементов земного покрова на основе временного ряда спутниковых изображений. Как показывают последние публикации в этой области [11], наиболее эффективным является ансамблевый нейросетевой метод классификации временных рядов данных. Особенностью предлагаемого метода является получение на выходе каждой из сетей не только значения конкретного класса, но и усредненной апостериорной вероятности классификации для каждого из классов.

III. Фильтрация изолированных пикселей и векторизация изображения. Для устранения шума предлагается использовать комбинацию логических и морфологических операций совместно с медианным фильтром [12]. Этот этап предполагает выделение на карте отдельных векторных объектов для последующей численной оценки площадей изменений в рамках отдельных административных единиц уровня района или области.

IV. Геопространственный анализ результатов, полученных на предыдущих стадиях, на основе векторной информации о типах земного покрова и границах административных районов и оценка изменений земного покрова за период 1990–2010 гг. Для семантического геопространственного анализа информации в задачах устойчивого развития на этом этапе используются статистические данные по отдельным регионам.

Оценка изменений земного покрова на основе метода глубинного обучения. В статье [13] приведены результаты картографирования земного покрова с использованием первых двух этапов предлагаемой в данной работе технологии глубинного обучения. В [13] показано, что полученная для территории Украины карта существенно точнее всех существующих аналогов. Третий этап, связанный с фильтрацией и векторизацией полученного изображения, детально описан в [14]. В соответствии с этими работами, наиболее эффективным является метод объектной классификации. Данный метод основан на анализе результатов попиксельной классификации, с учетом маски облачности и границ полигонов, с возможностью разделения их на несколько полей. В результате применения предлагаемого подхода построены векторные карты земного покрова для территории Украины за 1990, 2000 и 2010 гг. Основываясь на этих картах, сосредоточимся на верхнем (IV) уровне иерархической системы глубинного обучения, с помощью которого удастся решить задачу анализа изменений земного покрова.

В процессе классификации выделяются шесть основных классов, а именно: искусственные объекты, пахотные земли, лес, необрабатываемые земли, открытый грунт, вода.

На основе трех полученных карт строятся карты основных четырех переходов одних классов в другие за следующие промежутки времени: с 1990 по 2000, с 2000 по 2010 и общий переход с 1990 по 2010 гг.

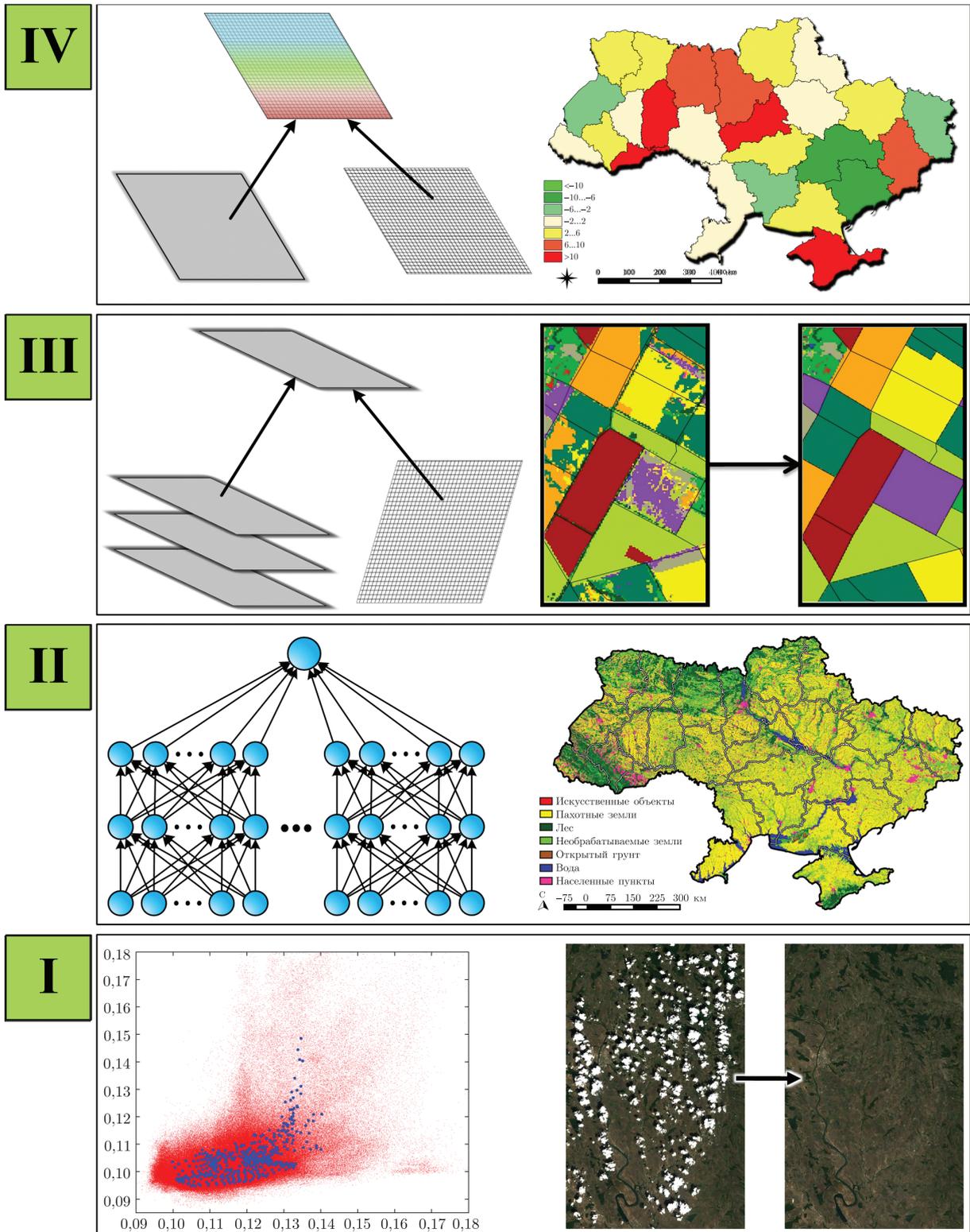


Рис. 1. Четырехуровневая иерархическая структура модели глубокого обучения для решения задачи оценки изменений земного покрова

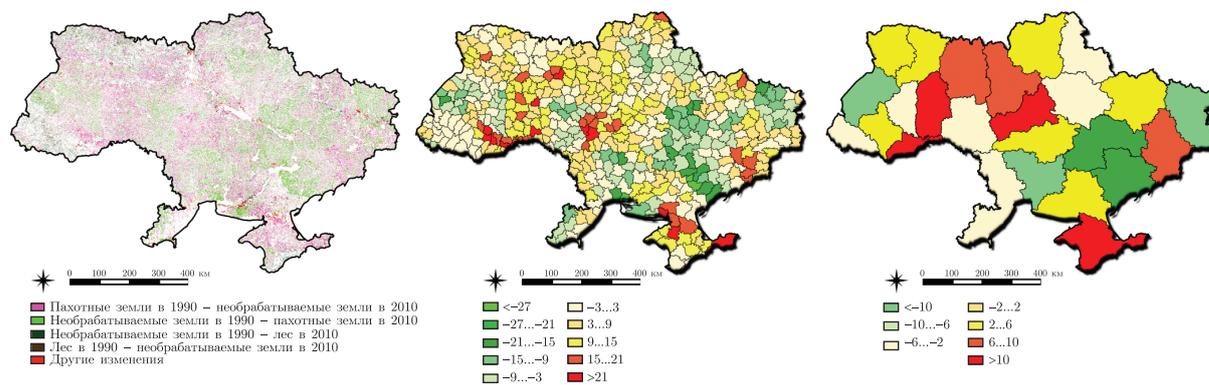


Рис. 2. Изменения земного покрова (землепользования) Украины с 1990 по 2010 г

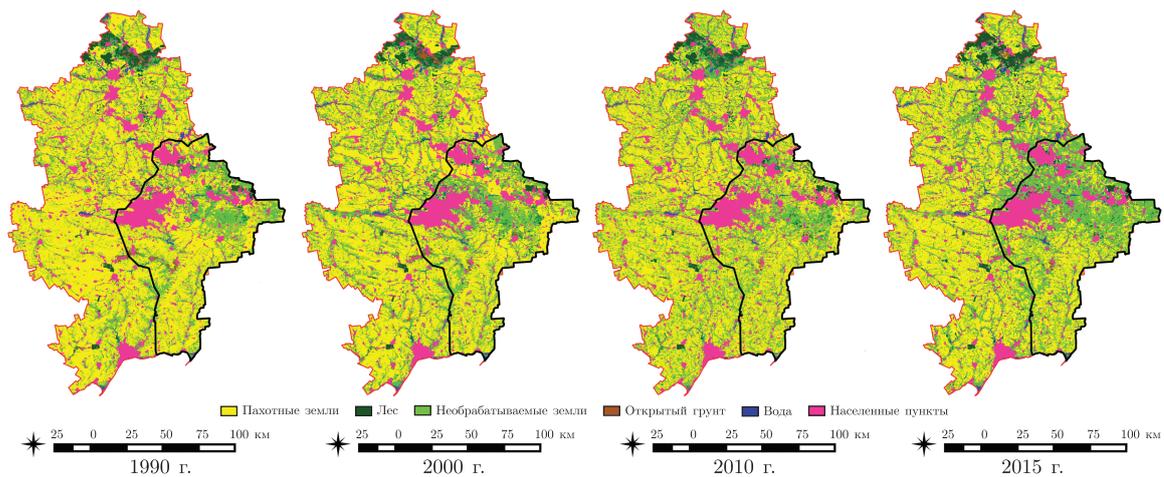


Рис. 4. Карты земного покрова для Донецкой области

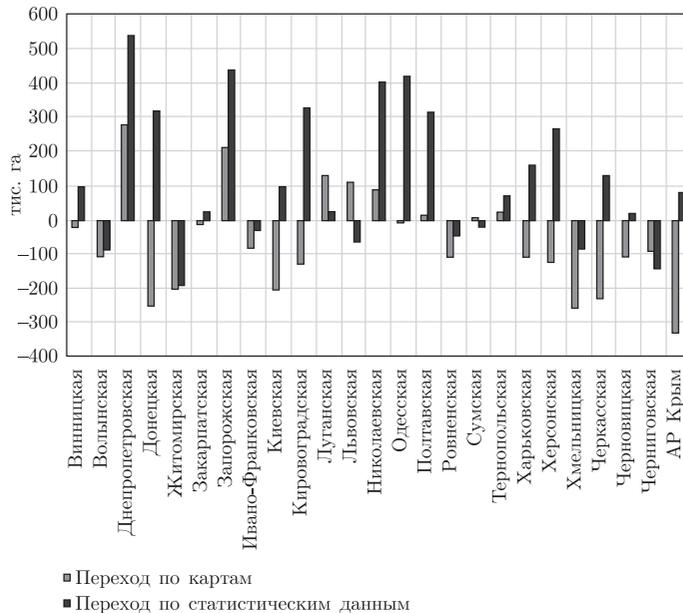


Рис. 3. Изменения площади пахотных земель за период 1990–2010 гг. по статистическим данным и по спутниковым картам классификации

В результате геопространственного анализа изменений земного покрова получены показатели на трех уровнях (пиксельном, в разрезе административных районов и областей). Графическое представление полученных результатов приведено на рис. 2 (см. вклейку). На растровой (пиксельной) карте (см. рис. 2, а) показаны основные переходы: лиловым цветом выделен переход пахотных земель в необрабатываемые, зеленым — необрабатываемых в пахотные. На рис. 2, б, в приводятся результаты обобщения по районам и областям, вычисленные по формуле

$$K(i) = \frac{S_1(i) - S_2(i)}{S_{obl}(i)} \cdot 100\%,$$

где $S_1(i)$ — площадь, которую занимает переход 1 (пахотные земли в необрабатываемые) в рамках i -й административной единицы; $S_2(i)$ — площадь 2-го перехода (необрабатываемые земли в пахотные); $S_{obl}(i)$ — площадь i -й административной единицы; $K(i)$ — процентный показатель прироста заброшенных земель в рамках i -й административной единицы по отношению к ее площади. Красным цветом выделены административные единицы, для которых увеличилась площадь необрабатываемых земель, а зеленым — те из них, для которых она уменьшилась.

Сопоставляя полученные результаты для областей Украины за период с 1990 по 2010 гг. с данными официальной статистики, можно увидеть определенные расхождения. Сравнительный анализ изменения площади пахотных земель для отдельных областей за период 1990–2010 гг. по статистическим данным и на основе спутниковых наблюдений приводится на рис. 3.

Полученные результаты свидетельствуют о том, что в северо-западных областях Украины заметной тенденцией является увеличение площадей необрабатываемых земель. Наибольшие площади заброшенных земель наблюдаются в Черновицкой, Хмельницкой, Черкасской областях и АР Крым. Несколько лучше ситуация в Житомирской и Киевской обла-

Таблица 1. Площадь пахотных и необрабатываемых земель для Донецкой и Луганской областей (1000 га)

| Класс | Донецкая область | | | | Луганская область | | | |
|------------------------|------------------|---------|---------|---------|-------------------|---------|---------|---------|
| | 2015 г. | 2010 г. | 2000 г. | 1990 г. | 2015 г. | 2010 г. | 2000 г. | 1990 г. |
| Пахотные земли | 1308,9 | 1522,2 | 1563,2 | 1695,5 | 1223,9 | 1396,4 | 1330,0 | 1257,4 |
| Необрабатываемые земли | 814,9 | 605,2 | 551,5 | 435,9 | 849,9 | 681,2 | 778,6 | 864,4 |

стях. И только во Львовской, Николаевской, Луганской, Запорожской и Днепропетровской областях наблюдается увеличение площадей пахотных земель.

Развивая полученные результаты, можно оценить изменения земного покрова и землепользования в Донецкой и Луганской областях, особенно на оккупированных территориях. На рис. 4 (см. вклейку) приведены карты классификации для Донецкой и Луганской областей, полученные на основе предложенного выше подхода. Черным цветом на рисунке показана линия разграничения.

Как видно из табл. 1, в Донецкой области площадь пахотных земель уменьшилась более чем на 20% по сравнению с 1990 г., причем значительные отрицательные изменения приходятся преимущественно на оккупированные территории и окрестности линии разграничения — 35%. Для Луганской области также наблюдается существенное сокращение пахотных площадей, примерно 2% для всей области и более 15% на оккупированных территориях. Очевидно, в результате боевых действий, которые происходят в восточных областях Украины, многие поля в Донецкой и Луганской областях в 2015 г. так и остались необработанными.

Таким образом, в связи с появлением в открытом доступе большого количества спутниковых данных возникает проблема обработки больших объемов данных и создания геоинформационных систем. Для ее решения развиваются методы науки о данных и машинного обучения. Авторы имеют значительные результаты в области методов машинного обучения и анализа геопространственных и спутниковых данных [15], которые валидированы не только для Украины, но и для различных регионов мира. На основе нейросетевых методов машинного обучения, разработанных специально для решения проблем больших массивов данных, а именно ансамблевых методов глубинного (deep learning) и активного машинного обучения (active machine learning), созданы автоматические информационные технологии распознавания типов земной поверхности по временным рядам спутниковых данных и решения других задач спутникового мониторинга. Эти задачи актуальны для поддержки принятия управленческих решений на региональном и государственном уровне.

Полученные результаты вносят весомый вклад в реализацию рабочих планов комитета наблюдения Земли GEO, а также его программ спутникового агромониторинга GEOGLAM и мониторинга рынков AMIS. Это открывает новые перспективы для европейской интеграции Украины и усиления роли Украины в глобальных международных программах комитета наблюдения Земли GEO и программы Европейского космического агентства Copernicus.

Цитированная литература

1. *Deng L., Yu D.* Deep learning: methods and applications // *Foundations and Trends in Signal Processing.* – 2014. – 7, No 3–4. – P. 197–387.
2. *Kussul N., Shelestov A., Basarab R., Skakun S., Kussul O., Lavreniuk M.* Geospatial intelligence and data fusion techniques for sustainable development problems // *ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer: Proc. 11th Int. Conf., ICTERI 2015, Lviv, Ukraine, May 14–16, 2015.* – Publ. CEUR-WS. Vol. 1356. – P. 196–203. – <http://ceur-ws.org/Vol-1356/>.

3. *Zgurovskiy M. Z., Pankratova H. D.* Системный анализ: проблемы, методология, приложения. – Киев: Наук. думка, 2005. – 743 с.
4. *Kussul N., Shelestov A., Skakun S., Kravchenko O.* High-performance intelligent computations for environmental and disaster monitoring // *Int. J. Information Technologies & Knowledge*. – 2009. – **3**. – P. 135–156.
5. *Kussul N., Shelestov A., Skakun S., Kravchenko O.* Data assimilation technique for flood monitoring and prediction // *Intern. J. Inform. Theories Appl.* – 2008. – **15**, No 1. – P. 76–84.
6. *Kussul N., Shelestov A., Skakun S.* Intelligent Computations for Flood Monitoring // *Advanced Research in Artificial Intelligence*. – Sofia: Institute of Information Theories and Applications FOI ITHEA, 2008. – P. 48–54. – (International Book Series “Information Science and Computing”; Book 2).
7. *Haykin S.* *Neural Networks and Learning Machines*. – 3rd ed. – Upper Saddle River, NJ: Prentice Hall, 2008. – 938 p.
8. *Rachkovskij D. A., Kussul E. M., Baidyk T. N.* Building a world model with structure-sensitive sparse binary distributed representations // *Biologically Inspired Cognitive Architectures*. – 2013. – **3**. – P. 64–86.
9. *Olofsson P., Foody G. M., Herold M., Stehman S. V., Woodcock C. E., Wulder M. A.* Good Practices for Estimating Area and Assessing Accuracy of Land Change // *Remote Sens. Environ.* – 2014. – **148**. – P. 42–57.
10. *Skakun S., Basarab R.* Reconstruction of Missing Data in Time-Series of Optical Satellite Images Using Self-Organizing Kohonen Maps // *J. Automat. Inform. Sci.* – 2014. – **46**, No 12. – P. 19–26.
11. *Kussul N., Skakun S., Shelestov A., Kravchenko O., Kussul O.* Crop classification in Ukraine using satellite optical and SAR images // *Models&Analyses*. – 2013. – **2**, No 2. – P. 118–128.
12. *Бакан Г. М., Куцсиль Н. Н.* Теоретико-множественная идентификация линейных объектов в классе размытых эллипсоидальных множеств // *Автоматика*. – 1990. – № 3. – С. 29–40.
13. *Лаврениук Н. С., Скакун С. В., Шелестов А. Ю., Яйлимов Б. Я., Янчевский С. Л., Яцук Д. Ю., Коттецкий А. М.* Классификация больших площадей земного покрова по ретроспективным спутниковым данным // *Кибернетика и системный анализ*. – 2016. – **52**, № 1. – С. 137–149.
14. *Kussul N., Lemoine G., Gallego J., Skakun S., Lavreniuk M.* Parcel based classification for agricultural mapping and monitoring using multi-temporal satellite image sequences // *The International Geoscience and Remote Sensing Symposium (IGARSS)*. – Milan, 2015. – P. 165–168. – DOI: 10.1109/IGARSS.2015.7325725.
15. *Куцсиль Н. Н., Скакун С. В., Шелестов А. Ю.* Анализ рисков чрезвычайных ситуаций на основе спутниковых данных. Модели и технологии. – Киев: Наук. думка, 2014. – 184 с.

References

1. *Deng L., Yu D.* *Foundations and Trends in Signal Processing*, 2014, **7**, No 3–4: 197–387.
2. *Kussul N., Shelestov A., Basarab R., Skakun S., Kussul O., Lavreniuk M.* Geospatial intelligence and data fusion techniques for sustainable development problems, ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer: Proc. 11th Int. Conf., ICTERI 2015, Lviv, Ukraine, May 14–16, 2015, Publ. CEUR-WS, Vol. 1356: 196–203, <http://ceur-ws.org/Vol-1356/>.
3. *Zgurovsky M. Z., Pankratova N. D.* *System analysis: problems, methodology, applications*, Kiev: Naukova Dumka, 2005 (in Russian).
4. *Kussul N., Shelestov A., Skakun S., Kravchenko O.* *Int. J. Information Technologies & Knowledge*, 2009, **3**: 135–156.
5. *Kussul N., Shelestov A., Skakun S., Kravchenko O.* *Intern. J. Inform. Theories Appl.*, 2008, **15**, No 1: 76–84.
6. *Kussul N., Shelestov A., Skakun S.* *Intelligent Computations for Flood Monitoring*, *Advanced Research in Artificial Intelligence*, International Book Series “Information Science and Computing”; Book 2, Sofia: Institute of Information Theories and Applications FOI ITHEA, 2008: 48–54.
7. *Haykin S.* *Neural Networks and Learning Machines*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 2008.
8. *Rachkovskij D. A., Kussul E. M., Baidyk T. N.* *Biologically Inspired Cognitive Architectures*, 2013, **3**: 64–86.
9. *Olofsson P., Foody G. M., Herold M., Stehman S. V., Woodcock C. E., Wulder M. A.* *Remote Sens. Environ.*, 2014, **148**: 42–57.
10. *Skakun S., Basarab R.* *J. Automat. Inform. Sci.*, 2014, **46**, No 12: 19–26.

11. Kussul N., Skakun S., Shelestov A., Kravchenko O., Kussul O. *Models&Analyses*, 2013, **2**, No 2: 118–128.
12. Bakan G. M., Kussul N. N. *Avtomatica*, 1990, No 3: 29–40 (in Russian).
13. Lavreniuk M. S., Skakun S. V., Shelestov A. Ju., Yalimov B. Ya., Yanchevskii S. L., Yaschuk D. Ju., Kostetsky O. M. *Cybernetics and System Analysis*, 2016, **52**, No 1: 137–149 (in Russian).
14. Kussul N., Lemoine G., Gallego J., Skakun S., Lavreniuk M. *The International Geoscience and Remote Sensing Symposium (IGARSS)*, Milan, 2015: 165–168, DOI: 10.1109/IGARSS.2015.7325725.
15. Kussul N., Skakun S. V., Shelestov A. Ju. *Disaster risk analysis based on satellite data. Models and Technologies*, Kiev: Naukova Dumka, 2014 (in Russian).

Поступило в редакцію 12.12.2015

Н. М. Куссуль, А. Ю. Шелестов, М. С. Лавренюк, І. М. Бутко

Інститут космічних досліджень НАН України та ДКА України, Київ

E-mail: nataliia.kussul@gmail.com, andrii.shelestov@gmail.com,
nick_93@ukr.net, butko.igor.68@gmail.com

Інформаційні технології глибинного машинного навчання для аналізу змін земного покриття

Запропоновано метод та інформаційну технологію геопросторового аналізу змін земного покриття на основі багаторічних супутникових спостережень. Оскільки ця проблема належить до класу задач з великим об'ємом даних (big data), для її розв'язання розроблено метод глибинного машинного навчання, що ґрунтується на ієрархічній нейромережевій моделі. Метод дає змогу розв'язувати широкий клас прикладних задач аналізу змін земного покриття та землекористування.

Ключові слова: big data, глибинне навчання, нейромережеві моделі, зміна земного покриття.

N. N. Kussul, A. Yu. Shelestov, M. S. Lavreniuk, I. N. Butko

Space Research Institute of the NAS of Ukraine and SSA of Ukraine, Kiev

E-mail: nataliia.kussul@gmail.com, andrii.shelestov@gmail.com,
nick_93@ukr.net, butko.igor.68@gmail.com

Information techniques of deep machine learning for the analysis of land cover changes

The paper proposes a method and an information technique for the geospatial analysis of land cover changes from long-term satellite observations. Since it is a big data problem, we propose a deep machine learning method for its solution, which is based on a hierarchical neural network model. The method allows solving the wide range of applied problems of the analysis of land cover changes and land use.

Keywords: big data, deep learning, neural network models, land cover changes.