

УДК 681.3.06

Об интеллектуальном управлении процессом анализа пространственных ситуаций

Беляков Станислав Леонидович

к.т.н, профессор каф. ИАСБ ИКТИБ ЮФУ

Россия, г. Таганрог, sbelyakov@sfedu.ru

Аннотация. Рассматривается задача формирования ситуационной осведомленности аналитика, исследующего пространственное описание поставленной проблемы. Ситуационная осведомленность формируется в процессе изучения картографических материалов, предоставляемых геоинформационным сервисом. Решение прикладной задачи осуществляется в условиях неопределенности относительно цели, способа и показателя качества конечного результата. В работе предлагается метод управления процессом поиска решения задачи, основанный на применении модели случайного леса.

Ключевые слова: Интеллектуальное управление, геоинформационные системы, принятие решений, ситуационная осведомленность.

About intelligent management of the process of analyzing spatial situations

Stanislav Belyakov

Candidate of Engineering Sciences, Professor at the Department of Information and Analytical Security Systems, ICTIS, SfedU

Russia, Taganrog, sbelyakov@sfedu.ru

Abstract. The problem of situation awareness of an analyst investigating a spatial situation is considered. Situation awareness is formed as the study of cartographic materials - maps, charts or plans. The solution of the applied problem is carried out in conditions of uncertainty regarding the goal, method and quality indicator of the final

result. The paper proposes a method for controlling the process of finding a solution to a problem, based on the application of a random forest model.

Keywords: Intelligent management, geographic information systems, decision making, situation awareness.

Необходимость в интерактивном анализе геоданных о сложившейся проблемной ситуации с целью прогнозирования ее развития и оценки возможных альтернатив решений возникает в различных областях бизнеса, производства и планирования. Находясь в контуре управления сложной системой, аналитики решают трудно формализуемые задачи, представленные пространственными данными и отношениями. Поиск решения осуществляется в интерактивном режиме. Формирование у аналитика ситуационной осведомленности является результатом осмысления визуальных образов карты рабочей области [1,2].

Задача управления анализом заключается в достижении требуемого уровня ситуационной осведомленности при наличии когнитивной нагрузки на аналитика. Когнитивная перегрузка сводит к минимуму эффективность интерактивного взаимодействия. В данной работе анализируется управление процессом анализа, основанное на знаниях. Перспективность подобного подхода определяется высокой степенью неопределенности и неполноты информации о диалоговом взаимодействии аналитика с геосервисом [3].

Рассмотрим особенности управления, построенного на применении контекстов [4]. Контекст представляет собой структуру данных, с помощью которой может быть сформирована рабочая область карты определенного смыслового содержания. Будем считать, что геосервис использует множество контекстов

$$C = \{c_1, c_2, \dots, c_M\},$$

каждый из которых может устанавливаться в сеансе произвольное число раз. Набор использованных в сеансе контекстов является прецедентом анализа,

$$\bar{c}_k = (\bar{c}_{1k}, \bar{c}_{2k}, \dots, \bar{c}_{n_k k}), \bar{c}_{ik} \in C,$$

а совокупность известных прецедентов представляет опыт проведения анализа, накопленный геосервисом

$$\bar{C} = \bigcup_k \bar{c}_k.$$

Обозначим через \bar{c}_w последовательность контекстов, которая используется в текущем сеансе анализа. Будем считать процесс анализа успешным, если аналитик достиг нужного уровня ситуационной осведомленности, использовав при этом минимально возможное число переходов от одного контекста к другому:

$$|\bar{c}_w| \rightarrow \min. \quad (1)$$

Чтобы достичь поставленной цели, следует учитывать следующее:

- уровень ситуационной осведомленности не поддается прямому оцениванию. Можно лишь утверждать, что он не убывает с увеличением длины последовательности \bar{c}_w ;

- выбор контекста, который будет использоваться в процессе анализа на следующем шаге, субъективен. Можно лишь констатировать, что удлинение \bar{c}_w происходит не хаотично, а рационально. Смысл и причина добавления нового контекста ясны только аналитику;

- опыт наблюдения процесса анализа (множество последовательностей \bar{C}) достоверно представляет возможные варианты направлений смыслового анализа.

Перечисленные факторы приводят к решению задачи управления (1) следующим образом: на каждом шаге анализа, изменяющего контекст, считать наилучшим тот, который максимально соответствует накопленному опыту. Это означает, что новый контекст $c_n \in C$ ранее был использован в самой короткой из известных последовательностей:

$$\begin{aligned} c_n \in \bar{c}_k, \bar{c}_k \in \bar{C}, \\ c_w \subset \bar{c}_k \ \& \ |\bar{c}_k| = \min_{\substack{\bar{c}_j \in \bar{C} \\ |\bar{c}_j| > |c_w|}} |\bar{c}_j|. \end{aligned} \quad (2)$$

Таким образом, в процессе анализа геосервис прогнозирует наилучший в смысле (1) контекст. Этот контекст следует рассматривать как осмысленное продолжение анализа [5,6].

Трудность нахождения $c_n \in C$ связана с недостатком информации об осмысленных последовательностях анализа. Накапливаемый геосервисом опыт лишь частично покрывает возможные реализации процесса. Разработать детерминированный алгоритм выбора контекста сложно из-за недоступности достоверных данных об уровне текущей ситуационной осведомленности аналитика. В качестве базового метода прогнозирования контекста будем использовать решающие деревья [7]. С их помощью можно накапливать знания о ходе анализа. Рассмотрим особенности применения данного метода.

Представим данные опыта из множества \bar{C} в виде подмножеств, содержащих последовательности длиной $1, 2, 3 \dots N$ элементов:

$$\bar{C} = \bigcup_i \bar{C}^{(i)},$$

где $\bar{C}^{(i)}$ - множество последовательностей длины i .

Последний элемент любой последовательности $c_n \in C$ будем считать значением прогноза. Тогда для каждого $\bar{C}^{(i)}$ может быть построено решающее дерево, в вершинах которого элементы текущей последовательности \bar{c}_w сравниваются с подмножествами контекстов, установленными при построении дерева. Результатом поиска в дереве является значение $c_n \in C$.

Поскольку последовательности имеют разную длину, опыт геосервиса представляется лесом из решающих деревьев. Используя известные механизмы выработки прогнозов с помощью леса, можно получить прогнозируемое значение $c_n \in C$.

Следует ожидать, что добавление новых экземпляров последовательностей в \bar{C} улучшит качество прогнозирования [7]. Вместе с тем дополнительная возможность повысить качество заключается в структурировании опыта. Выделим

в каждой последовательности $c_n \in C$ две части:

$$c_n = c_n^l \cup c_n^o, \quad (3)$$

С содержательной точки зрения, первая часть (c_n^l) представляет собой «долгосрочный» компонент анализа. В него входят контексты, использованные на всем протяжении сеанса анализа за исключением некоторого завершающего интервала времени τ . Содержание компонента c_n^l – это экспериментальные данные, отражающие рост ситуационной осведомленности пользователя в процессе анализа. Они несут в себе информацию о том, что в анализе временных рядов называют «level, intercept» [8]. Компонент c_n^o отражает «оперативный» этап анализа. В него входят контексты, использованные при завершении анализа. Эти контексты характеризуют тренд [8] процесса, поскольку стали следствием предшествующих размышлений аналитика. Критерием отбора контекстов в c_n^o является их попадание во временной интервал τ , составляющий несколько минут [9]. Важность выделения компонента c_n^o в том, что входящие в него контексты с большой вероятностью приводят к завершению сеанса анализа и свидетельствуют о достижении необходимой ситуационной осведомленности.

Представление (3) позволяет усовершенствовать прогнозирование предварительным определением оперативного компонента c_n^o . Наблюдаемая последовательность контекстов \bar{c}_w используется для выявления с помощью решающих деревьев наиболее перспективного направления дальнейшего анализа. Поскольку контексты внутри c_n^o упорядочены, аналитику предлагается следовать этой последовательности. Если он этого не делает, то прогнозируется новый тренд с учетом выбранного контекста.

Практическая полезность предложенного метода определяется точностью прогнозирования контекста. Для экспериментальной оценки этой величины был использован корпоративный геосервис, включающий около 30 контекстов. Для 200

реализаций сеансов анализа с помощью библиотеки scikit-learn [10] была построена модель случайного леса. В качестве переменных использовались только идентификаторы контекстов. На тестовой выборке ошибка составила 18%. Данное значение следует считать удовлетворительным, поскольку более высокая точность предсказания, с нашей точки зрения, свидетельствовала бы скорее о переобучении модели. Перспективным путем повышения точности может стать добавление переменных, характеризующих пространственные и временные границы контекстов.

Библиографический список

1. Longley P. A., Goodchild M., Maguire D. J., Rhind D. W. *Geographic Information Systems and Sciences*, 3rd ed. Wiley, 2011, 1580 p.
2. Shashi S., Hui X. *Encyclopedia of GIS*. New York: SpringerScience+Buisiness Media, LLC, 2008, 720 p.
3. Goodchild M.F. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 2007, 69(4), 211-221.
4. Dey A., Abowd G. Towards a better understanding of context and context-awareness, CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness, 2000, pp. 304–307.
5. Belyakov S., Bozhenyuk A., Rozenberg I. The intuitive cartographic representation in decision-making. In: *World Scientific Proceeding Series on Computer Engineering and Information Science*, 2016, vol.10, pp. 13-18.
6. Belyakov S., Belyakova M., Savelyeva M., Rozenberg I. The Synthesis of Reliable Solutions of the Logistics Problems Using Geographic Information Systems. In: *10th International Conference on Application of Information and Communication Technologies (AICT)*, IEEE Press, New York (2016), pp. 371-375.
7. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning*. Springer-Verlag New York, 2009, 480 p.
8. Kane M.J., Price N., Scotch M. et al. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics*,

2015, 15, pp. 276-288.

9. Gibson J.J A Theory of Direct Visual Perception. In: J. Royce, W. Rozenboom (eds.). The Psychology of Knowing. Gordon & Breach, New York, 1972, 270 p.

УДК 681.3.06

Оценка достоверности результатов поиска информации

Котов Эдуард Михайлович

Старший преподаватель каф. ИАСБ ИКТИБ ЮФУ

Россия, г. Таганрог, emkotov@sfedu.ru

Аннотация. Решая задачу ранжирования результатов информационного поиска, всегда возникает вопрос об эффективности функции ранжирования, насколько нам подходят традиционные меры оценки информационного поиска действительно ли исследуемый метод лучше другого и если производилась оценка более чем для одного запроса, как все это отразится на достоверности поисковых результатов. Эти темы исследует автор статьи.

Ключевые слова: Информационный поиск, ранжирование, меры оценки.

Evaluating the reliability of information search results

Eduard Kotov

Senior Lecturer at the Department of Information and Analytical Security Systems,

ICTIS, SfedU

Russia, Taganrog, emkotov@sfedu.ru

Abstract. When solving the problem of ranking information search results, there is always a question about the effectiveness of the ranking function, whether traditional measures of evaluating information search are suitable for us, whether the method under study is better than another, and if more than one query was evaluated, how all this will