

Стронская Наталия Тарасовна

Трансформация библиотечно-информационной отрасли в контексте актуализации биобиблиографического наследия в цифровую эпоху

Освещен вопрос алгоритмизации биобиблиографических данных происходящей посредством считывания наличествующей культурологической составляющей, анализа некоего дискурса, в котором бытует персоналия. Утверждается, что исследования исторического и культурологического аспекта биобиблиографических данных является интегрированным не только вследствие многоплановости и широты научного предмета, но и дифференцированности бытования индивидуума. Его перспективой представляется разработка посредством просчитывания оптимальных для исследования биобиблиографических данных алгоритмов его работы.

Ключевые слова: биобиблиографические данные; биобиблиографическое наследие; биобиблиографический поиск; SVM; метод опорных векторов; алгоритмизация.

Natalia T. Stronskaja

Transformation of the Library and Information Industry in the Context of Updating the Biobibliographic Heritage in the Digital Age

The article deals with the issue of algorithmization of biobibliographic data occurring through reading the existing cultural component, the analysis of a discourse in which there is a person. It is argued that the study of historical and cultural aspects of biobibliographic data is integrated not only due to the diversity and breadth of the scientific subject, but also the differentiation of existence of the individual. His perspective is the development of algorithms optimal for the study of biobibliographic data by calculating its operation.

Keywords: biobibliographic data; biobibliographic heritage; biobibliographic search; SVM; support vector machine; algorithms.

Биобиблиографические данные представляют собой некий особый вид данных, носящих социально-детерминированный характер своей природы: с одной стороны, они социально стратифицированы, с другой – культурологически обусловлены. Последнее делает их особенно самобытными: в некотором роде такие данные можно воспринимать в качестве информации, сходной по своей природе с фольклорной. Это связано с тем, что такие данные не просто обусловлены и продуцированы культурной эпохой, само человеческое бытие (существование) в ней можно рассматривать как составляющую культурной полисистемы.

Таким образом, говоря об особенностях актуализации биобиблиографических данных мы подразумеваем целый пласт культуры, вовлеченный посредством новых ее проявлений (форм), материализованных через информационные технологии. В таком разрезе само создание искусственного интеллекта представляется скорее культурно-гносеологической задачей, нежели строго эмпирической.

Если говорить конкретнее, то, безусловно, актуализация культурной составляющей исторических исследований в обличье биобиблиографического сегмента информации закономерно представлена таким явлением как алгоритмизация. Реализация последней возможна посредством череды отдельных алгоритмов, совокупная деятельность которых повлечет достижение неких исследовательских целей.

В частности, в глобальной сети Интернет содержится огромное количество разнообразной информации. Однако самое трудное в работе с сетью Интернет в контексте автоматизации биобиблиографического поиска является определение корректных ключей,

по которым должен производиться поиск и ранжирование конечного результата. Поскольку Интернет не имеет четкой централизованной структуры, развивается хаотично, и в мире появляются все новые и новые серверы, вопрос поиска, в том числе биобиблиографической информации становится очень актуальным.

Поисковые системы значительно облегчают работу в Интернет и помогают нам быстро найти нужную информацию в огромном массиве серверов Интернет (WWW, FTP и др.). Во всемирной паутине находится несколько тысяч поисковых систем, среди которых есть как те, что уже хорошо зарекомендовали себя, так и менее известные. В последнее время, в связи с возрастающим объемом био- и библиографических данных, рост которых будет только увеличиваться в математической прогрессии, актуальным стал вопрос автоматизации поиска и составления таблиц биобиблиографических данных по интересующим нас рубрикам. Для решения этой задачи целесообразно создание класса специализированных поисковых машин, которые будут обучаемым инструментом сетевого биобиблиографического поиска.

Как известно, каждая из поисковых машин имеет свои преимущества и недостатки, которые определяются принципом их работы, удобством использования, ее оформлением, сложностью языка запросов, наличием различных расширенных функций, скоростью работы. Выбор поискового алгоритма или их сочетаний для конкретного применения определяется целью поиска, характером искомой информации и желаемым форматом данных.

Целью данной работы является представление одного из самых перспективных среди используемых при анализе больших массивов данных методов – метода опорных векторов (SVM-Support vector machines). Он адаптирован к определению связей между заданными и искомыми ключами по задаваемому алгоритму, хотя для применения его на практике, необходима мощная информационная база.

Основная идея данного метода: отразить входной вектор различных признаков через нелинейное отражение. SVM показал лучшие классификационные результаты по сравнению с параметрическими методами и таким популярным и широко используемым непараметрическим, как метод нейронных сетей, считается одним из самых точных. В отличие от последнего, SVM имеет достаточно привлекательные свойства, предоставляет единое решение, характеризующееся глобальным минимумом оптимизированных функциональных и безальтернативных решений. Более того, SVM в значительной степени полагается на метод эвристики и имеет более гибкую структуру.

SVM приобретает все большую популярность благодаря своим особенностям и способности обобщать широкий круг проблем. В частности, этот метод имеет два основных преимущества:

- а) он рассматривает линейные «нештатные» ситуации, расширяет возможности модели, по поиску неизвестных переменных для прогнозирования;
- б) он принимает принцип структурной минимизации рисков [3].

Таким образом, этот метод может предложить хорошие поисковые возможности, но может не быть выбранным для практического применения по определению потенциальных связей между ключом и массивом биобиблиографических данных (по крайней мере, отдельно от других методов) для составления сводной таблицы по запросу [5]. Продемонстрировав актуальность практического использования метода опорных векторов для осуществления ранжирования и поиска биобиблиографической информации, теперь можем дать формулировку Лагранжа [4].

В линейном случае справедливы неравенства для всех n -начальной выборки:

$$\chi^T_t w + b \geq 1 - \xi_i \text{ для } y_i = 1;$$

$$\chi_t^T w + b \leq -1 + \xi_i \text{ для } y_i = -1;$$

$$\xi_i \geq 0;$$

которые могут быть объединены в два ограничения:

$$y_i(\chi_t^T w + b) \geq 1 - \xi_i; \quad (1)$$

$$\xi_i \geq 0; \quad (2)$$

Основная идея SVM состоит в том, чтобы найти такое распределение гиперплоскости библиографических данных, что соответствует максимально возможным запасам между точками разных классов.

Итак, SVM основан на очень немногих ограничительных предположениях и может обнаружить те эффекты, которые выпускаются из виду многими другими методами. Он способен и адаптирован к установлению точных результатов классификации в других областях, и может стать вариантом выбора для установления более точного библиографического поиска. Однако для того, чтобы создать практически ценную методологию нужно объединить SVM с большим набором библиографических данных и обратиться к альтернативным формулировкам метода опорных векторов, которые лучше подходят для обработки больших наборов данных. В целом, у нас есть ценный инструмент для осуществления библиографического поиска, который способен отвечать требованиям первоначальных поисковых условий.

Таким образом, алгоритмизация библиографических данных происходит именно посредством считывания наличествующей культурологической составляющей, анализа некоего дискурса, в котором бытует персоналия. Подобное исследование, бесспорно, носит интегрированный характер не только вследствие многоплановости и широты научного предмета, но и дифференцированности бытования индивидуума. Его перспективой представляется разработка SVM посредством просчитывания оптимальных для исследования библиографических данных алгоритмов его работы.

1. *Boser, B. E.* A training algorithm for optimal margin classifier / B. E. Boser, I. M. Guyon, V. N. Vapnik // Proc. 5th Annual ACM Workshop on Computational Learning Theory. – ACM Press, 1992. – P. 144-152.

2. *Vapnik, V.* Support-vector networks / V. Vapnik, C. Cortes // Machine Learning. – 1995. – 20 (3). – P. 273-297.

3. *Chen, S.* Modeling default risk with support vector machines / S. Chen, W. K. Härdle, R. A. Moro // Quantitative Finance. – 2011. – 11 (1). – P. 135-154.

4. *Hardle, W.* Predicting Bankruptcy with Support Vector Machines / W. Hardle, R. Moro, D. Schafer. – Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Germany, 2008. – P. 11-14.

5. *Kaya, M. E.* An analysis of support vector machines for credit risk modeling / M. E. Kaya, F. Gurgen, N. Okay // Applications of Data Mining in E-Business and Finance, Vol. 177, Frontiers in Artificial Intelligence and Applications, Edited by C. Soares, Y. Peng, J. Meng, T. Washio and Z.-H. Zhou. – ACM Press, 2008. – P. 25-33.

Туралина Неонила Альфредовна
Сушкова Юлия Николаевна

Дискурсивное пространство библиотеки: инновационный подход к проблеме

В статье раскрывается характеристика инновационного подхода и дискурсивного пространства библиотечной организации, анализируется специфика, сущность и функции