

ТЕХНОЛОГИИ СЕМАНТИЧЕСКОЙ ОБРАБОТКИ ИНФОРМАЦИИ В УЧЕБНОМ ПРОЦЕССЕ



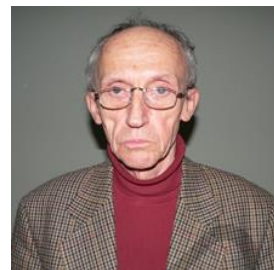
Л.Ю. Шилин

Декан факультета информационных технологий и управления БГУИР, доктор технических наук, профессор



А.А. Навроцкий

Заведующий кафедрой информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент



Л.С. Стригалев

Старший преподаватель кафедры информационных технологий автоматизированных систем БГУИР

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: dekfitu@bsuir.by, navrotsky@bsuir.by, orion@bsuir.by

Abstract. Обсуждаются методологические вопросы применения технологий семантической обработки информации в учебном процессе.

Сложность, высокая динамичность развития и стоимость современных технологий порождают многочисленную проблематику, предъявляя повышенные требования к образовательным технологиям, особенно в IT-сфере, которая имеет не только всевозрастающую долю ВВП в мировой экономике, но и в силу своего конвергентного характера порождает дополнительные специфические проблемы, в том числе и методологические проблемы подготовки IT-специалистов.

Важным и одновременно сложным «срезом» образовательной технологии становится адекватный мониторинг качества работы студента, необходимый для его оптимальной проводки в образовательном пространстве. В современных условиях академических свобод и использования традиционных технологий, включая «облачные» [1–4], для эффективной реализации образовательных траекторий явно недостаточно. К тому же необходим и современный конструктивный методологический аппарат, поскольку традиционная образовательная парадигма не только не отображает системную сложность образовательных технологий и объектов, но и не находит адекватного места для систем и средств компьютерного интеллекта, которые в ряде областей достигли уровня «интеллекта» среднего человека.

Объекты образовательной среды (обучаемый, обучающий и любая обучающая организация) могут быть описаны четверкой: система, структура, цель, технология, с тремя взаимосвязанными уровнями целеполагания: генетический, неосознанный и осознанный [2, 3].

Обучение человека осуществляется на неосознанном и осознанном уровне целеполагания как информационно-энергетическая составляющая метаболизма, которая имеет три технологических уровня: синтаксический, семантический и прагматический. Синтаксический уровень (восприятие, преобразование, передача, хранение информации) в обучении носит обеспечивающий характер; в технических системах это по большей части хорошо разработанный инфраструктурный уровень.

Важнейшим уровнем в обучении является неосознанный уровень (условный и каузальный рефлексы; ментальность, привычка). На этом уровне в процессе обучения формируется общая культура и профессиональные компетенции. Осознанная целенаправленность, на

уровне которой формируется образовательная траектория, играет активную роль в формировании личности и совместно с внешней средой воздействует на структуру неосознанной целенаправленности. Пространство свободы человека, который обучается практически в течение всей жизни (но фундамент интеллекта закладывается в когнитивном возрасте, до 17-18 лет) определяется генетическим и неосознанным уровнями, а также состоянием внешней среды.

Так как процесс обучения реализуется на семантическом технологическом уровне, то учебный материал и результаты обучения носят структурированный характер и, следовательно, имеют конечную колмогоровскую сложность, что позволяет автоматизировать оценку качества процесса обучения. Колмогоровская сложность характеризует индивидуальные свойства объекта, чем она меньше, тем выше структурированность объекта. Структурированный объект с вероятностью близкой к единице порождает эргодические (типичные) реализации. Определенным аналогом колмогоровской сложности (но, по множеству реализаций) является шенноновская избыточность.

Неструктурированный объект, имеющий нулевую избыточность (максимальную колмогоровскую сложность), не имеет семантики. Наличие избыточности свидетельствует о существовании закономерности (структуры), порожденной либо интеллектуальной системой, либо объектом с известными или непознанными свойствами. Очевидно, что система (естественная или искусственная) способная обнаруживать, анализировать и «работать» с семантикой является интеллектуальной, в особенности, если она способна еще и принимать решения. Подобного рода системы становятся все более востребованными, что напрямую связано с понятием «интернет вещей» (Internet of Things, IoT). По мнению компании Cisco Интернет вещей это момент времени, когда количество материальных объектов, подключенных к Интернету, превысило количество людей, пользующихся Интернет; что произошло в промежутке между 2008 и 2009 годами.

Особую роль в деле семантической обработки информации играют технологии Data Mining. Понятие Data Mining, появилось в 1978 году; популярность приобрело примерно в первой половине 1990-х годов. Особенно актуальны технологии Data Mining в настоящее время, что обусловлено всевозрастающими объемами неструктурированной информации, которую порождает не только человек, традиционные технологические и научно-исследовательские средства и системы, но и многочисленные мобильные устройства, количество которых значительно превышает население Земли. В результате возник технологический разрыв между возможностями и потребностями в обработке семантических свойств информации, который потребовал развития адекватных средств искусственного интеллекта, реализующих семантическую обработку. Возможности средств искусственного интеллекта на данный момент в ряде случаев соответствуют уровню среднестатистического человека.

О роли и назначении Data Mining в обработке информации достаточно красноречиво говорят варианты перевода этого термина: добыча данных, "раскопка" данных, извлечение знаний, извлечение "зерен знаний" из гор данных, "промывание" данных, информационная проходка данных, интеллектуальный анализ данных и т. д. Сказанное, по существу, позиционирует и многочисленные сферы применения Data Mining.

Технологии Data Mining, которые по существу входят в состав методов анализа больших данных (Big Data [5]), развивалась на базе прикладной статистики (это стартовая позиция данной технологии), теории искусственного интеллекта и машинного обучения, теории баз данных и других областей. Следует заметить, что в ряде источников технологии методов Big Data и Data Mining пересекаются, например, машинное обучение, распознавание образов и др.

Технологии Data Mining можно классифицировать по многочисленным признакам. По типу обрабатываемых данных Data Mining подразделяются:

- Text Mining — технологии поиска и семантического анализа текста;
- Web Mining — интеллектуальный анализ данных в Internet;
- Call Mining — «добыча звонков», технология распознавания речи и ее анализ;

- Audio Mining — извлечение данных из аудиозаписей;
- Video Mining — извлечение данных из видеозаписей.

Data Mining имеет неограниченную, постоянно расширяющуюся сферу применения, однако наибольший эффект, который, как отмечается в ряде источников, может достигать 1000% характерен для коммерческих предприятий, что в конкурентной борьбе делает неотвратимым применение данной технологии. Кроме коммерческой сферы технологии Data Mining находят применение в производстве, телекоммуникации, государственном управлении, научных исследованиях, медицине, геологоразведке и т.д.

Одним из перспективных направлений Data Mining является технологии Educational Data Mining (EDM), которые ориентированы на исследования данных, используемых в образовательных целях, для анализа и принятия решений в сфере образования. EDM (первая конференция прошла в 2008 году в Монреале), в виду сложности образовательных технологий, имеют определенную специфику, которая помимо стандартных методов Data Mining, предполагает использование и других специфичных методов, например, методов психометрии.

В современных условиях применение технологий Data Mining в сфере образования является не только перспективным, но и необходимым как по экономическим, так и по стратегическим соображениям. Для достижения максимального эффекта эти технологии должны внедряться как в основные, так и обеспечивающие процессы учреждения образования, а также в процессы мониторинга системы качества этого учреждения. Главным направлением применения технологии EDM является оптимальная проводка студента в образовательном пространстве с максимальным использованием генетического потенциала студента.

Литература

[1]. Батура М.П. Дистанционное образование: концепция, технологии, контент, сервисы / М.П. Батура, Б.В. Никульшин, В.Ю. Цветков // Дистанционное обучение – образовательная среда XXI века: Материалы VII Междунар. научн.-метод. Конференции, 1-2 декаб. 2011 г. – Минск: БГУИР, 2011 – С.7-12.

[2]. Strigalev L.S, German O.V. Methodological aspects of the IT-specialists training // Информационные технологии и системы 2011: Материалы Международной конференции, Минск, БГУИР, 2011 - С.199, 200.

[3]. Стригалева Л.С. Слабоструктурированные аспекты технологии дистанционного обучения. // Дистанционное обучение – образовательная среда XXI века: Материалы VI Междунар. научн.-метод. конференции, 22-23 нояб. 2007 г. Минск : БГУИР. 2007. С.230-232.

[4]. Никульшин Б.В. О создании частного образовательного облака / Б.В. Никульшин, В.Е., Проволоцкий, Е.М. Димидюк, Л.С. Стригалева / Информационные технологии и системы 2013 (ИТС 2013): Междунар. научн.-метод. конференции, 23 октября 2013. Минск: БГУИР, 2013. С. 304-305.

[5]. Шилин Л.Ю. Технология больших данных как стратегическое направление / Л.Ю. Шилин, Навроцкий А.А., Герман О. В., Л.С. Стригалева // BIG DATA and Predictive Analytics. Минск: БГУИР, 2016. С. 271-273.