

Правительство Республики Беларусь опирается на политику импортозамещения с 1990-х годов и регулярно издает постановления, содержащие перечень и объемы импортозамещающей продукции (в первую очередь потребительских товаров), которые, как ожидается, будут производиться отечественными компаниями.

В Беларуси также стимулировалось производство импортозамещающей продукции и резидентами СЭЗ. В 2008 г. ряд поправок был внесен в указы Президента, касающиеся деятельности СЭЗ. В результате была установлена упрощенная процедура получения от республиканских органов управления одобрения перечня импортозамещающих товаров, производимых и реализуемых на территории страны резидентами СЭЗ.

Принимая во внимание сходство льгот, предоставляемых всем СЭЗ, существует два фактора, которые администрации СЭЗ могут использовать для привлечения новых инвесторов. К ним относятся ориентированный на интересы бизнеса подход при управлении и обеспечение привлекательной инфраструктуры. В то же время следует подумать о реализации различных стратегий развития в СЭЗ. Это уменьшит порочную конкуренцию между СЭЗ в пределах одной страны, что является следствием предложения практически одинаковых налоговых льгот [2].

Список использованных источников

1. Лесняк О. В. Свободные экономические зоны как способ привлечения иностранных инвестиций // Молодой ученый. – 2017. – №45. – С. 189–190.

2. Беларусь: Сравнительное исследование промышленных парков и особых экономических зон, ноябрь 2018 года. Европейский Банк Реконструкции и Развития. – С. 34–36.

УДК 004.89+004.6

Н.В. Пацей, Д.В. Шиман, А.С. Наркевич, И.Г. Сухорукова
Белорусский государственный технологический университет

МЕТОДЫ ОЧИСТКИ И ПОДГОТОВКИ ИНФОРМАЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА

В настоящее время количество анализируемой информации значительно увеличилось, кроме того по самой своей природе и содержанию она стала более разнообразной и обширной. Современные бизнес-требования

это уже не простой поиск или статистическая обработка, а более сложный процесс интеллектуального анализа [1–2].

Интеллектуальный анализ данных основан на построении подходящей модели, которую можно использовать для обработки, выявления и создания необходимой информации. Независимо от формы и структуры источника данных, информация структурируется и организуется в соответствии с определенным форматом.

В зависимости от источника данных особенно важно выбрать правильный способ построения и преобразования этой информации, каким бы ни был метод окончательного анализа. Этот шаг связан со сложным процессом сбора и очистки информации.

Таблица 1 – Методы очистки и преобразования информации

Очистка данных	
Заполнение пропущенных данных	Удаление/игнорирование пропусков
	Заполнение пропуска средним значением
	Использование глобальных нулей
	Повторение результата последнего наблюдения
	Использование наиболее вероятного значения
Идентификация выбросов и сглаживание шумов	корреляционно-регрессионный метод
	Удаление выбросов на основе частот появления значений
	Интервальный метод
	Строгие формальные правила
	Контрольные числа
Фильтрация данных	Использование машинных словарей
	Анализ строк
Преобразование данных	
Сглаживание методом спектральной обработки	Спектральный анализ
	Авторегрессионные методы
Агрегированные операции	
Нормализация (масштабирование) данных	min-max нормализация
	z-нормализация
	масштабирование
Дискретизация	Биллинг
	Гистограмма
	Кластерный анализ.
Удаление избыточности данных и оптимизация	
Редукция данных	Уменьшение размерности: вейвлет-преобразования; метод главных компонент (РСА); выделение подмножества атрибутов
	Компактное представление данных: регрессия, сэмплинг, OLAP куб
	Сжатие данных
Снижение размерности	Метод главных компонент
	Метод факторного анализа

Цель обработки данных низкого качества теряет всякий смысл независимо от того, какие алгоритмы и технологии используются. Если анализируемые данные не соответствуют определенным критериям качества, то их предварительная обработка является обязательным этапом для обеспечения адекватного результата [2].

Основные методы и алгоритмы очистки и преобразования данных представлены в таблице 1.

В идеале очистка и пред обработка данных должна улучшить качество модели (но, как показала практика, не во всех случаях). Кроме того, процесс очистки требует больших временных затрат на устранение даже простых ошибок (удаление выбросов, дубликатов и т.п.). Структурно процессы схожи, но для разных доменов и наборов данных отличаются. Таким образом, возникает задача просмотра списка функций очистки/обработки и проверки, применяются ли они и как параметризовать функции.

Был предложен алгоритм итеративной структурированной обработки, который может корректно переобучать модель очистки/преобразования данных и предоставляет на выбор набор оптимизаций. Обработке подвергается небольшое подмножество данных для получения модели очистки, которая затем используется уже на полном наборе. В качестве инструмента разработки использовался Python.

Для проверки устойчивости модели очистки к ошибкам данных было выполнено тестирование, при котором внедрялись различные ошибки данных, чтобы проверить устойчивость моделей интеллектуального анализа [3]. Как показывают предварительные эксперименты применение алгоритма структурированной обработки позволяет повысить абсолютную точность классификации от 3% до 9% в зависимости от качества первоначального набора, а также дает трехкратное ускорение процесса очистки.

Список использованных источников

1. Van den Broeck, J., Cunningham, S. A., Eeckels, R., Herbst, K. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS medicine*, 2(10), с.267.

2. Когаловский М.Р., Методы интеграции данных в информационных системах [Электронный ресурс] – <http://www.ipr-ras.ru/articles/kogalov10-05.pdf>

3. Панченко О.Л., Пацей Н.В. Оценка работы текстового классификатора на основе методов машинного обучения // доклады XVIII Международной конференции Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2019), Минск, 21 ноября 2019 г. – Минск : ОИПИ НАН Беларуси, 2019. – С. 185–189.