

УДК 004.42:519.254

МАТЕМАТИЧЕСКИЕ МЕТОДЫ И ПРОГРАММНЫЕ ИНСТРУМЕНТЫ ДЛЯ ОЦЕНКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО СОСТОЯНИЯ КЕМЕРОВСКОЙ ОБЛАСТИ

Овчинников Е.Н., студент гр. ПИБ-151, IV курс

Научный руководитель: Пимонов А.Г., д.т.н., профессор

Кузбасский государственный технический университет имени Т.Ф. Горбачева
г. Кемерово

По последним данным сводных отчетов о развитии экономики и социальной сферы Кемеровской области развитие социально-экономического состояния показывает положительную динамику. Однако официальные рейтинги, охватывающие все множество субъектов Российской Федерации, не отражают в должной степени конкурентоспособность [1] Кемеровской области с субъектами, внешние условия функционирования которых сопоставимы. К этим условиям можно отнести климат, цены на энергоресурсы, затраты на логистику и т. п.

Качество анализа и риски неверной оценки напрямую зависят от выбранных критериев, качества выборки и используемых методов [2] и инструментов. В данной работе рассмотрены существующие методы и инструменты для анализа статистических данных [3], результаты применения которых дают достаточно наглядное представление о социально-экономическом положении Кемеровской области на период актуальности исходных данных.

Для получения качественной оценки показателя социально-экономического состояния региона [4, 5] необходимо выделить ключевые показатели, проанализировав которые можно сделать выводы о текущем состоянии и по возможности спрогнозировать изменения.

Для определения наиболее важных соотношений и связей внутри системы можно провести разведочный анализ согласованности показателей этой системы. Для его проведения используются различные корреляционные методы, позволяющие получить как параметрические, так и непараметрические показатели корреляции. Для комплексного анализа можно использовать несколько показателей. Например, линейный коэффициент корреляции (коэффициент корреляции Пирсона) и коэффициент ранговой корреляции (Кендалла/Спирмена). Результатами вычислений будут корреляционные матрицы, содержащие коэффициенты корреляции для всех возможных пар показателей.

Получить структуру, имеющую интерпретации относительно значимости, связей и диспропорций показателей можно в результате проведения анализа специфики показателей. В терминах описательной статистики специфичность показателей отражает долю отрицательных результатов, которые правильно идентифицированы как таковые. Проведение анализа энтропийным методом [6] дает предварительное энтропийное отображение показате-

лей. Визуализировать итоги позволяет получаемая математическая конструкция – «Фазовый портрет свойств».

Важно определить параметры совокупности показателей: наличие явных лидеров, имеющихся подсистем и подобия состояний, присутствие резко выделяющихся по величине показателей. Для этого можно провести анализ состояния объектов, который возможно реализовать с применением энтропийного метода. В результате анализа получается «Фазовый портрет системы», а структура данной математической конструкции позволяет получить необходимые параметры.

Для проведения аудита результатов энтропийного анализа, а также получения дополнительных результатов, помогающих в некоторой степени визуализировать принадлежность исследуемых объектов к определенной группе, можно использовать метод кластеризации. Результатом вычислений по данному метода является диаграмма, на которой точки, изображающие объекты, маркируются в соответствии с принадлежностью к тому или иному кластеру. Наиболее известным и проверенным методом кластеризации является метод k -средних.

С целью аудита результатов метода k -средних и энтропийного анализа, а также более наглядного представления многомерных данных целесообразно произвести анализ многообразия. Результатом данного анализа будет диаграмма, на которой изображены исследуемые объекты, и которая максимально наглядно изображает структуру всей исследуемой системы.

Во всех видах проводимого анализа получаемые результаты напрямую зависят от представительности набора объектов и значимости исследуемых показателей.

После определения перечня методов для анализа социально-экономического состояния субъекта следует рассмотреть доступные программные реализации, позволяющие составить схему анализа данных и провести требующиеся вычисления.

При выборе языка программирования для реализации всех представленных методов встает вопрос не столько быстродействия самого программного обеспечения, сколько скорости его разработки и простоты сопровождения. Достаточную инфраструктуру предоставляет набирающий популярность в сфере анализа данных и машинного обучения язык Python. В силу открытости исходного кода как самого языка, так и всех используемых в нем стандартных и авторских библиотек именно на нем следует остановить выбор.

В качестве среды разработки в сфере Data Mining и машинного обучения наибольшее распространение получили Jupyter, Spyder и Orange. Выбор конкретной среды разработки зависит от личных предпочтений, так как все они имеют последние версии интерпретаторов языка Python (Anaconda, Iron Python, IPython) и совместимы с большинством пакетов анализа данных.

Наиболее известные и стабильно поддерживаемые сообществом разработчиков библиотеки для анализа данных – NumPy, Pandas, Matplotlib, SciKit-learn, SciPy, Keras, TensorFlow. Среди представленных библиотек, некоторые

обладают обширным набором средств визуализации данных (Matplotlib, Pandas), оптимизированными алгоритмами работы с матрицами и математическими вычислениями (NumPy, Pandas), большим выбором готовых алгоритмов машинного обучения и возможностью параллельных вычислений на GPU (Keras, TensorFlow) и оптимизированными средствами обработки числовых последовательностей (SciPy). Но наиболее простым в использовании и при этом имеющий весь достаточный набор методов является пакет SciKit-learn. В данном пакете присутствуют готовые реализации методов анализа согласованности показателей – «Pearson Correlation» и «Spearman Correlation», специфичности показателей – «Features Rating», состояния объектов – «Entropy Transform», кластеризации объектов методом k -средних – « k -Means» и анализа многообразия – «Manifold Learning».

Анализ, проведенный с использованием выбранных программных инструментов и математических методов, позволит достичь необходимой наглядности в отношении свойств Кемеровской области и минимизировать трудозатраты на проверку результатов, их интерпретацию и дальнейшее сопровождение разработанных алгоритмов.

Список литературы

1. Фридман, Ю.А. Конкурентные позиции региона в условиях инновационного развития экономики / Ю.А. Фридман, Г.Н. Речко, А.Г. Пимонов // Регион: экономика и социология. – 2016. – №4(92). – С. 218–236.
2. Дороганов, В.С. Методы статистического анализа и нейросетевые технологии для прогнозирования показателей качества металлургического кокса / В.С. Дороганов, А.Г. Пимонов // Вестник Кемеровского государственного университета. – 2014. – №4, Т. 3. – С. 123–129.
3. Тынкевич, М.А. Статистический анализ данных на компьютере / М.А. Тынкевич, А.Г. Пимонов, А.М. Вайнгауз; Кузбас. гос. техн. ун-т им. Т.Ф. Горбачева.– Кемерово, 2013.– 124 с.
4. Фридман, Ю.А. Оценка и анализ влияния драйверов роста на конкурентоспособность региона / Ю.А. Фридман, Г.Н. Речко, А.Г. Пимонов, Ю.Ш. Блам // Региональная экономика. Юг России. – 2016. – №2 (12). – С. 25–35.
5. Фридман, Ю.А. Оценка влияния уровня инновационности развития сырьевой отрасли на конкурентоспособность региона / Ю.А. Фридман, Г.Н. Речко, А.Г. Пимонов // Вестник Кузбасского государственного технического университета. – 2015. – №5. – С. 167–178.
6. Арнаутов, Р.С. Система поддержки принятия решений для управления портфелем ценных бумаг на основе энтропийных мер риска / Р.С. Арнаутов, А.Г. Пимонов, К.Э. Рейзенбук // Вестник Кузбасского государственного технического университета. – 2015. – №6. – С. 169–175.