

# Отчет о проверке на заимствования №1

Автор: МФТИ [admin@phystech.edu](mailto:admin@phystech.edu) / ID: 211

Проверяющий: ([admin@phystech.edu](mailto:admin@phystech.edu) / ID: 211)

Организация: Московский физико-технический институт

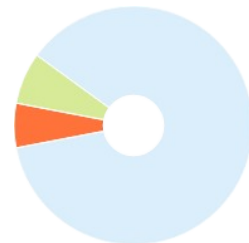
Отчет предоставлен сервисом «Антиплагиат» - <http://mipt.antiplagiat.ru>

## ИНФОРМАЦИЯ О ДОКУМЕНТЕ

№ документа: 1105  
Начало загрузки: 27.06.2018 12:23:27  
Длительность загрузки: 00:00:56  
Имя исходного файла:  
Usmanova2018CCM\_PLS.pdf  
Размер текста: 845 кБ  
Символов в тексте: 31192  
Слов в тексте: 4184  
Число предложений: 236  
Способ извлечения текста: OCR

## ИНФОРМАЦИЯ ОБ ОТЧЕТЕ

Последний готовый отчет (ред.)  
Начало проверки: 27.06.2018 12:24:25  
Длительность проверки: 00:00:13  
Комментарии: не указано  
Модули поиска: Модуль поиска "МФТИ", Цитирование, Кольцо вузов, Модуль поиска общепотребительных выражений, Модуль поиска перефразирований Интернет, Модуль поиска перефразирований eLIBRARY.RU, Модуль поиска Интернет, Коллекция eLIBRARY.RU, Коллекция РГБ, Сводная коллекция ЭБС



ЗАИМСТВОВАНИЯ	ЦИТИРОВАНИЯ	ОРИГИНАЛЬНОСТЬ
5,91%	7,04%	87,05%

Заимствования — доля всех найденных текстовых пересечений, за исключением тех, которые система отнесла к цитированиям, по отношению к общему объему документа.  
Цитирования — доля текстовых пересечений, которые не являются авторскими, но система посчитала их использование корректным, по отношению к общему объему документа. Сюда относятся оформленные по ГОСТу цитаты; общепотребительные выражения; фрагменты текста, найденные в источниках из коллекций нормативно-правовой документации.  
Текстовое пересечение — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.  
Источник — документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.  
Оригинальность — доля фрагментов текста проверяемого документа, не обнаруженных ни в одном источнике, по которым шла проверка, по отношению к общему объему документа.  
Заимствования, цитирования и оригинальность являются отдельными показателями и в сумме дают 100%, что соответствует всему тексту проверяемого документа.  
Обращаем Ваше внимание, что система находит текстовые пересечения проверяемого документа с проиндексированными в системе текстовыми источниками. При этом система является вспомогательным инструментом, определение корректности и правомерности заимствований или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего.

№	Доля в отчете	Доля в тексте	Источник	Ссылка	Актуален на	Модуль поиска	Блоков в отчете	Блоков в тексте
[01]	1,03%	1,96%	Read eBooks online   World Heritage En...	<a href="http://community.worldheritage.org">http://community.worldheritage.org</a>	26 Июл 2017	Модуль поиска Интернет	3	7
[02]	0,43%	1,58%	Singular spectrum analysis	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	раньше 2011	Модуль поиска Интернет	2	8
[03]	0,49%	1,25%	Forecasting of EEG Data by Singular Spec...	<a href="http://elibrary.ru">http://elibrary.ru</a>	раньше 2011	Коллекция eLIBRARY.RU	2	6
[04]	0%	1,1%	<a href="http://orca.cf.ac.uk/54926/1/U585315.pdf">http://orca.cf.ac.uk/54926/1/U585315.pdf</a>	<a href="http://orca.cf.ac.uk">http://orca.cf.ac.uk</a>	18 Июл 2017	Модуль поиска Интернет	0	6
[05]	0,81%	1,05%	Causality studied in reconstructed state ...	<a href="http://arxiv.org">http://arxiv.org</a>	26 Мар 2017	Модуль поиска Интернет	2	2
[06]	0,03%	1%	2016_ИЭИТУС_ПОВТАС_090401_МД_ПА...	не указано	20 Июн 2016	Кольцо вузов	1	4
[07]	0%	0,96%	Александров, Фёдор Игоревич диссерт...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	5
[08]	0%	0,89%	Анализ сингулярного спектра для филь...	<a href="http://elibrary.ru">http://elibrary.ru</a>	28 Авг 2014	Коллекция eLIBRARY.RU	0	5
[09]	0,63%	0,87%	Robust Non-linear Wiener-Granger Caus...	<a href="http://arxiv.org">http://arxiv.org</a>	26 Мар 2017	Модуль поиска Интернет	1	2
[10]	0%	0,77%	Говорунов, Денис Александрович дисс...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	3
[11]	0%	0,68%	Факультет экономики/71-ФРИН Султан...	не указано	26 Мая 2014	Кольцо вузов	0	3
[12]	0%	0,68%	Modeling daily realized futures volatility ...	<a href="http://elibrary.ru">http://elibrary.ru</a>	раньше 2011	Коллекция eLIBRARY.RU	0	3
[13]	0%	0,63%	Does Trading Volume Contain Informatio...	<a href="http://elibrary.ru">http://elibrary.ru</a>	раньше 2011	Коллекция eLIBRARY.RU	0	2
[14]	0,59%	0,59%	Построение кросс-корреляционных за...	<a href="http://elibrary.ru">http://elibrary.ru</a>	02 Янв 2018	Модуль поиска перефразирований eLIBRARY.RU	1	1
[15]	0%	0,54%	Сорокин, Анатолий Александрович ди...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	3
[16]	0,23%	0,52%	Фоменко, Мария Викторовна диссериа...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	1	2
[17]	0%	0,52%	Математические методы обработки н...	<a href="http://biblioclub.ru">http://biblioclub.ru</a>	20 Апр 2016	Сводная коллекция ЭБС	0	2
[18]	0,51%	0,51%	FDI, local Financial Markets, employmen...	<a href="https://mp.ra.uni-muenchen.de">https://mp.ra.uni-muenchen.de</a>	01 Июл 2017	Модуль поиска Интернет	2	2
[19]	0%	0,51%	Лукин, Глеб Владимирович диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	2
[20]	0%	0,5%	Усевич, Константин Дмитриевич диссе...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	2

[21]	0%	0,5%	Удумян, Давид Каджикович диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	2
[22]	0,48%	0,48%	ДиссертацияДата размещения: 02.10.2...	<a href="https://guu.ru">https://guu.ru</a>	14 Дек 2016	Модуль поиска Интернет	1	1
[23]	0%	0,47%	Флоринский, Игорь Васильевич диссе...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	2
[24]	0%	0,45%	kulikov_v_i_scenarnye-metody-prognozi...	не указано	28 Апр 2017	Кольцо вузов	0	1
[25]	0%	0,45%	kulikov_v_i_sposobny-li-makroekonomic...	не указано	17 Мая 2018	Кольцо вузов	0	1
[26]	0%	0,45%	golovin_o_a_effektivnost-transgranichny...	не указано	23 Мая 2018	Кольцо вузов	0	1
[27]	0%	0,45%	ershov_k_a_ekonomicheskie-krizisy-i-fon...	не указано	29 Мая 2018	Кольцо вузов	0	1
[28]	0%	0,44%	46739	<a href="http://e.lanbook.com">http://e.lanbook.com</a>	10 Мар 2016	Сводная коллекция ЭБС	0	2
[29]	0%	0,43%	Соломаха, Илья Викторович диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	2
[30]	0%	0,41%	Нгуен Минь Хай диссертация ... кандид...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	2
[31]	0,37%	0,37%	Investigation of Causal Relationship betw...	<a href="http://elibrary.ru">http://elibrary.ru</a>	раньше 2011	Коллекция eLIBRARY.RU	1	1
[32]	0%	0,35%	Заверняев, Константин Валерьевич ди...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	20 Янв 2010	Коллекция РГБ	0	1
[33]	0%	0,33%	rahimova_n_sh_vliyanie-makroekonomi...	не указано	10 Мая 2017	Кольцо вузов	0	1
[34]	0%	0,33%	apryshko_a_e_prognozirovanie-dinamik...	не указано	25 Мая 2018	Кольцо вузов	0	1
[35]	0%	0,33%	Факультет мировой экономики и мир...	не указано	04 Июн 2012	Кольцо вузов	0	1
[36]	0%	0,33%	Testing for fiscal competition among Fre...	<a href="http://elibrary.ru">http://elibrary.ru</a>	25 Авг 2014	Коллекция eLIBRARY.RU	0	1
[37]	0%	0,33%	Optimal operating arrangements in the r...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[38]	0%	0,33%	Conservative incompetence, Labour resp...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[39]	0%	0,33%	Energy and the use of conservation tillag...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[40]	0%	0,33%	Aggregation and the role of energy in the...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[41]	0%	0,33%	Metal prices and the business cycle.	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[42]	0%	0,33%	The effect of the price of gold on its prod...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[43]	0%	0,33%	Causal Flow between Human Well-being...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[44]	0%	0,33%	The Relationship between Bond Returns...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[45]	0%	0,33%	Investing in Research and Education vers...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[46]	0%	0,33%	Explaining Post-War Cinema Attendance...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1
[47]	0%	0,33%	The minimum complexity of endogenous...	<a href="http://elibrary.ru">http://elibrary.ru</a>	раньше 2011	Коллекция eLIBRARY.RU	0	1
[48]	0%	0,33%	Серпионов, Виктор Юрьевич диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	1
[49]	0%	0,33%	Замураев, Константин Александрович...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	1
[50]	0%	0,33%	Рожков, Андрей Григорьевич диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	1
[51]	0%	0,33%	Наговицын, Юрий Анатольевич диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	1
[52]	0%	0,33%	Назруллаева, Евгения Юрьевна диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	07 Мар 2012	Коллекция РГБ	0	1
[53]	0%	0,33%	Ильин, Евгений Александрович диссертаци...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	1
[54]	0%	0,33%	Пономаренко, Владимир Иванович си...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	1
[55]	0%	0,33%	210272	<a href="http://biblioclub.ru">http://biblioclub.ru</a>	18 Апр 2016	Сводная коллекция ЭБС	0	1
[56]	0%	0,33%	211420	<a href="http://biblioclub.ru">http://biblioclub.ru</a>	18 Апр 2016	Сводная коллекция ЭБС	0	1
[57]	0%	0,33%	211419	<a href="http://biblioclub.ru">http://biblioclub.ru</a>	раньше 2011	Сводная коллекция ЭБС	0	1
[58]	0%	0,33%	257592	<a href="http://biblioclub.ru">http://biblioclub.ru</a>	19 Апр 2016	Сводная коллекция ЭБС	0	1
[59]	0,32%	0,32%	137423	<a href="http://biblioclub.ru">http://biblioclub.ru</a>	15 Апр 2016	Сводная коллекция ЭБС	1	1
[60]	0%	0,3%	Characteristics of Earnings-Leading Versu...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	1

[61]	0%	0,3%	Diversification Issues in Real Estate Inves..	<a href="http://elibrary.ru">http://elibrary.ru</a>	раньше 2011	Коллекция eLIBRARY.RU	0	2
[62]	0%	0,3%	Interactions Between Property and Equi...	<a href="http://elibrary.ru">http://elibrary.ru</a>	22 Авг 2014	Коллекция eLIBRARY.RU	0	2
[63]	0%	0,3%	Аскеров, Эльдар Нариманович диссер...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	2
[64]	0%	0,29%	Факультет мировой экономики и мир...	не указано	21 Мая 2012	Кольцо вузов	0	1
[65]	0%	0,19%	Смирнов, Дмитрий Алексеевич диссер...	<a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	раньше 2011	Коллекция РГБ	0	1
[66]	6,97%	0%	не указано	не указано	раньше 2011	Цитирование	2	2
[67]	0,07%	0%	не указано	не указано	раньше 2011	Модуль поиска общеупотребительных выражений	1	1

Текст документа

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
«МОСКОВСКИЙ ФИЗИКО—ТЕХНИЧЕСКИЙ ИНСТИТУТ  
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ) >>  
ФИЗТЕХ-ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ  
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ  
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»>>  
Усманова Карина Равилевна  
Модели обнаружения зависимостей ВО временных рядах В  
задачах построения ПРОГНОСТИЧЭСКИХ моделей  
03.03.01 — Прикладные математика И физика  
ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАВОТА  
(ВАКАЛАВРСКАЯ ДИССЕРТАЦИЯ)  
Научный руководитель:  
Д. ср.—М. Н. Стризов ВАДИМ Викторович  
Москва  
2018 г.

Аннотация

При прогнозировании сложноорганизованных временных рядов, зависящих от экзогенных факторов и имеющих множественную периодичность, требуется решить задачу выявления свя— занных рядов. Предполагается, что добавление этих рядов в модель повышает качество прогно— за. Статистическая значимость повышения качества прогноза выявляется в с помощью теста Гренджера. В данной работе для обнаружения связей между временными рядами предлагается использовать метод сходящегося перекрестного отображения. При таком подходе два временных ряда считаются зависимыми, если существует отображение окрестности фазовой траектории из одного лагированного пространства в другое. Также при обнаружении причинно—следственных связей между временными рядами ставится задача обнаружения оптимального лагированного подпространства. Решение этой задачи продемонстрировано на двух парах рядов: потребления электроэнергии и температура, объема железнодорожных перевозок нефти и цена на нефть.

Ключевые слова: временные ряды; прогнозирование; тест гренджеро; сжодничесн перекрестные отображения; оптимальное ,дозированное подпространство

Содержание

1 Введение

2 Постановка задачи прогнозирования

3 Обнаружение связей временных рядов

4 Алгоритм многомерной гусеницы (MSSA-L)

5 Метод проекций на латентные структуры

6 Вычислительный эксперимент

6.1 Сгенерированные данные .....

6.2 Эксперимент на данных потребления электроэнергии и температуры .....

6.3 Перебор подпространств .....

6.4 Построение прогноза .....

6.5 Эксперимент на данных об объеме перевозок нефти .....

7 Заключение

10

11

11

13

18

18

19

22

1 Введение

Работа посвящена обнаружению причинно—следственных связей между разнородными временными рядами. Примеры зависимых разнородных временных рядов: связь эконометрических временных рядов, связь показателей ЭКГ и пульса (<http://smartlab.ws/component/content/article?id=60>)

Если прогноз временного ряда  $X$  строится с использованием временных рядов  $y_1, \dots, y_t$ , то установление связей ряда  $X$  с  $y_1, \dots, y_t$ , может повысить качество прогноза и упростить прогно— стическую модель. Если установлено, что ряд  $X$  не зависит от ряда  $y$ , то  $y$  можно исключить из прогностической модели. В данной работе для обнаружения зависимостей между рядами в работе анализируются два подхода: тест Гренджера [1, 2] и метод сходящегося перекрестного отображения (convergent Cross mapping, CCM **14**) [3, 4].

В основе теста Гренджера лежит следующий подход. Считаем, что ряд  $X$  зависит от ряда  $y$  (или следует из ряда  $y$ ), если использование истории ряда  $y$  при построении прогностической модели статистически значимо повышает качество прогноза ряда  $X$ , [1, 2]. Тест Гренджера позволяет установить связи между рядами и основан на сравнении качества прогноза, в котором используется история только прогнозируемого ряда, и прогноза, который дополнительно использует историю других рядов. Если улучшение качества прогноза подтверждается статистически, то говорят, что прогнозируемый ряд связан с использовавшимися во втором прогнозе рядами. Тест Гренджера применим к стационарным временным рядам, поэтому в случае нестационарных рядов их необходимо продифференцировать перед проведением теста Гренджера. Тест Гренджера используется в различных задачах, в которых необходимо исследовать взаимосвязь между развивающимися во времени процессами [5, 6].

Недостатком теста Гренджера является то, что при используемом в нем подходе невозможно точно определить структуру зависимости рядов. Например, два ряда могут следовать из третьего, но при отсутствии информации о третьем ряде тест Гренджера установит причинно—следственную связь между первым и вторым рядом, хотя она отсутствует. Проблема точного определения структуры зависимости рядов рассмотрена в работе [7].

В случае, когда тест Гренджера неприменим или не может обнаружить связь между рядами, применяется метод сходящегося перекрестного отображения (convergent Cross mapping, CCM). Этот метод основан на оценке того, насколько хорошо один ряд может быть восстановлен и использован вторым. Считается, что ряд  $X$  точно восстанавливается по ряду  $y$ , только если ряд  $y$  влияет на ряд  $X$ . Метод CCM основан на сравнении ближайших соседей в траекторном пространстве ряда  $X$ , полученных с помощью ряда  $X$  и с помощью ряда  $y$ . Другими словами, проверяется, насколько точно моменты времени, соответствующие ближайшим соседям вектора  $y_t$ , определяют ближайших соседей вектора  $x_t$ . [3, 4].

При построении линейной прогностической модели по временному ряду строится траекторная матрица, играющая роль матрицы объектов. Ответами являются значения ряда в последующие моменты времени. Иногда размерность траекторного пространства очень велика, и, как следствие, прогностическая модель становится неустойчивой. В этом случае необходимо производить отбор признаков [8, 9]. Метод проекций на латентные структуры (partial least squares, PLS) отбирает наи—

более значимые признаки и строит новые признаки как их линейные комбинации [10, 11]. Таким образом, PLS находит подпространство траекторного пространства, проекция в которое наилучшим образом приближает исходный ряд. Снижение размерности может применяться при изучении связей между рядами. Проекция в траекторное подпространство позволяет более детально изучить связь между главными компонентами рядов и найти подпространство, в котором наблюдается связь

между рядами.

В данной работе для построения прогноза одного временного ряда по нескольким использу—  
ется алгоритм многомерной гусеницы (multivariate singular spectrum analysis, MSSA—L) [12]. Этот  
алгоритм является обобщением на 1\m1ногомерный случай алгоритма анализа спектральных компо—  
нен (smgular spectrum analysis, SSA) [13, 14, 15]. Метод SSA ОСНОВан На разложении временного  
ряда в сумму интерпретируемых компонент. Он делится на четыре основных шага: запись ряда в  
виде траекторной матрицы, ее сингулярное разложение, группировка компонент полученных при  
сингулярном разложении, по каждой сгруппированной матрице восстанавливается временной ряд.  
Таким образом исходный временной ряд представляется в виде суд/путы временных рядов. Метод  
SSA применяется в таких задачах, как выявления трендов во временных рядах [16], подавления  
шума во временных рядах [17], прогнозирование временных рядов [18, 19].

2 НОСТановКа ЗЕЪДЕЪЧИ ПРОГНОЗИРОВЕЪНИЯ

Поставим задачу прогноза многомерного временного ряда.

(1) (s) T<sub>1, ..., n, ..., n</sub>

Обозначим X : [X, . . . ,X \* ] 1 заданныи 3—мерный временной ряд. Построим матрицу плана

из СеГь/ТЭНТОВ РЯД&3

\$5" ... \$533,

: X0+(~\_1). (1)

513(8) . . . 337:1

Пусть X<sub>n</sub> : [51:79), . . . 3:55]] T 1 значение ряда X в момент времени n. Построим прогноз X ряда X в  
точке Xn. Проделаем это k: раз для различных обучающих выборок ХЁЩШ : X<sub>i</sub>+(n+;~\_1), i : 0, . . . , (k: — 1).

Получим k: прогнозов X : [X... X<sub>nn</sub>, . . . X<sub>n</sub>+;<sub>i,~\_1</sub> ряда X в точках X... XП", . . . X<sub>n</sub>+;<sub>i,~\_1</sub>.

НрОФНОСТquCKaH 1V10d6ЛЬ PIA/188T ВИД

Xt+1 = Ё(®,Xг,Xг—1, - - -7X1Е—b— |—2>7

A A A A

W : argmmS(w,X,Xn,Xn+1, . . . Xn+k\_1) : S(W,X,X),

где функция потерь

k—I k—I

^ ^ 1 A 1

S<W, X, X) : Z E(Xn—I—27 Xn—I—z) : Z E<\$7('L—)]—Z7 \$\$\$—ix)

2:0 2:0

3 Обнаружение связей временных рядов

Тест Гренджера. В качестве базового метода установления связей предлагается использовать  
статистический тест Гренджера. Ниже приведен алгоритм теста Гренджера для проверки наличия

связи Двух временных рядов. Требуется проверить, зависит ли ряд X от ряда y. Выдвинем гипотезу  
о независимости ряда X от ряда y и проверим ее. Делаем это следующим образом.

1. Строим прогноз ряда X без использования ряда y и находим значение функции потерь

n

Sx = E C (E131,371):

i=1

где n \* длина тестовой выборки.

Функцию Д(:.13;Ё) выбираем в зависимости от распределения ошибок прогноза на тестовой  
выборке.

2. СТРОИМ ПРОГНОЗ РЯД& X с ИСПОЛЬЗОВ&НИ6М ряд& y. Вь1ЧИСЛЯ6М для него ЗН&Ч6НИ6 ФУНКЦИИ  
потерь

ЗХУ : Zn: Д (2171, 51%).

i=1

3. Рассмотрим статистику

N — 2k: Sx — ху

k Sху’

где N \* длина обучающей выборки, k: \* размерность регрессионной модели. Статистика Т

Т(х7 у) :

имеет распределение F (k, N — 2k) (распределение Фишера с параметрами (k, N — 2%)).

4. Если ряд X не зависит от ряда y, то значения Sx и Sху будут близки, а статистика Т (X, y) \*  
незначима. Поэтому в случае больших значений статистики Т (X, y) отвергаем гипотезу о неза—

висимости ряда X от у. Выберем некоторое критические значение t статистики T (X, у). Тогда

критерий зависимости ряда X от ряда у выглядит следующим образом:

Из T(X, у) > t следует, что ряд X зависит от ряда у

5. Аналогично проверим зависимость ряда X от восстановленного (с помощью алгоритма MSSA—

L) ряда у. Для этого используем статистику

N—2k SX— x9

T ^ :

(жу) k Sxy

Для более подробного изучения связи между временными рядами X и у вычисляем кроссй

корреляционную функцию "Уд/(Ь)

1w) : E [(x, — их) (№ — „ул

2 2

охоу

7

2 i выборочная дисперсия. Если h\*

где E i математическое ожидание, ,и i выборочное среднее, 0

соответствует 1х4аксимальному значению кросс—корреляции, то говорят, что ряд у сдвинут на h\*

относительно X. Заметим, что если ряд X сдвинут на h1 относительно ряда у, а ряд у сдвинут на №

относительно ряда Z. То ряд X сдвинут на A;; : h1 + № относительно ряда Z.

Пусть прогноз ряда X строится с использованием истории ряда у и пусть с помощью вычисления

кросс—корреляции рядов X и у получено, что ряд X отстает от ряда у на h отсчетов времени. Тогда

ИСПОЛЬЗОВ&НИ6 ПРИ ПРОГНОЗЭ РЯД& у, СДВИНУТОГО на h отсчетов Н&3&Д, МОЖЭТ ПОВЫСИТЬ KaqтВО

прогноза.

Метод сходящегося перекрестного отображения, CCM. Опишем, как строится прогноз ря—

да у : [ш, . . . , уN] С ПОМОЩЬЮ ряда X : [31:1, . . . , :1: N]. Построим матрицу Ганкеля ряда х.

371 372 37L—1 37L

372 373 37L £17L+1

NX : 7

CCM—ьн £17N—L+2 --- CCM—1 £17N

где L \* ширина окна, длина истории ряда, используемая при нахождении главных компонент. Ана—

логично строим матрицу Ну. Обозначим 75—10 строку матрицы NX через Xt+L\_1 соответственно.

Тогда матрица NX принимает вид

XL

NX: £̂ , Xi:[37i—Ь+17---7372'—175£̂ij]7 i=L7---7N-

XN

Заметим, что все вектора X L, . . . ,хм принадлежат Ь—Ь/ерному траекторному пространству MX ряда

х. Аналогично вводим уt, t : L, . . . , N, лежащие в траекторном пространстве Му ряда у. Выберем

момент t E [L, N] и найдем k: ближайших соседей вектора х,, в MX. Обозначим их индексы через

t1, . . . , 75k. Тогда ближайшие соседи хг i это строки матрицы NX С номерами (t1 — L + 1), . . . , (75k — L + 1):

XtiZ[mti—L—l—la'uamti—lamti]: Z21,...,k7

Прогноз Qt строится следующим образом:

k

gt : E ушу," где tr индексы ближайших соседей х,,

i=1

2

„ХТ? \_ ХИ

Wz' и, : exp —

22' “i,

Аналогично строится прогноз ряда X С использованием ряда Y.

„ХТ? \_ XtL+1H2

ПОК&Ж61\4, как ОПИС&ННЬ1Й ПОДХОД ВЛОЖНО ПРИЫШОНЯЮТСЯ ДЛЯ ОБНАРУЖОНИЯ ВаВ/лС/нВЮСТ/И ВЛОЖДУ рН-

дами. Пусть выбран момент времени 75\* и вектор Xfi : [33,5\*\_b+1, . . . , :zjt\*\_1, жд]. И пусть ХИ, . . . ,th \*

ближайшие соседи вектора хдк. Тогда вектора уд, уtl, . . . ,ум i строки матрицы Ну, соответствую—

щие индексам t1, . . . , 75k . Тогда, если вектора уд, уtl, . . . ,уд, расположены в Му достаточно близко,

то утверждается, что ряд X зависит от ряда X.

Введем меру близости  $уд, utl, \dots, уд$ , следующим образом:

$k$   
 $dist(x) 1$   
 $S = - d' t : - * - .$   
 $(... Сижy, 1s <x> „ ; \th хьь$

Если  $S(x, y)$  меньше некоторого порога  $З$ , то ряд  $y$  зависит от ряда  $X$ .

Заметим, что можно рассматривать ближайших соседей не во всем траекторном пространстве  $MX$  и  $My$ , а только в некотором его подпространстве, натянутом на первые главные компоненты. Пусть сингулярное разложение матрицы  $HX$  имеет вид

$HX : UxAxVx$   
  
Пусть  $7; *$  некоторый набор индексов компонент ряда  $X$ . Построим проекцию ряда  $X$  на под—пространство, натянутое на компоненты с номерами из  $7; .$  Обозначим это подпространство  $MT; -$  Заменим  $B$  матрице  $A$ ; элементы, находящиеся в строках с индексами, не из  $7; ,$  нулями. Обозначим полученную матрицу  $Ax$ . Тогда проекция ряда  $X$  в подпространство, натянутое на компоненты с индексами из  $7;$  задается траекторной матрицей  
 $PTX : UXA XVX.$

Аналогично по некоторому набору  $Ty$  строится подпространство  $MTy$  и траекторная матрица  $Pty$ . Далее предлагается искать ближайших соседей не в полных траекторных пространствах  $MX$  и  $My$ , задающихся траекторными матрицами  $Ha$ , и  $Hу$  соответственно, а в подпространствах № $7;$  и  $MTy$ , задающихся матрицами  $PTX$  и  $PTY-$  Рассмотрев различные подпространства, можно выбрать то, которое будет наилучшим образом описывать исследуемый временной ряд и будет иметь минимальную размерность. Перебор различ—ных подпространств также позволяет установить, между какими именно компонентами рядов  $X$  и  $y$  существует зависимость.

Зависимость рядов в выбранных подпространствах устанавливается аналогично зависимости в полных пространствах  $MX$  и  $My$ . Пусть  $7;$  и  $Ty i$  наборы индексов главных компонент рядов  $X$  и  $y$  соответственно. Тогда  $PTX$  и  $Pty i$  траекторные матрицы проекции рядов  $X$  и  $y$  в подпро—странства, натянутые на  $7;$  и  $Ty$  соответственно. Для фиксированного  $75^*$  находим  $k$ : ближайших соседей  $XI, \dots, th$  и соответствующие им  $yt^*, utl, \dots, y, \dots$ . Здесь  $X$ , и  $yt 7$  строки матриц  $PTX$  и  $Pty$  соответственно.

Будем перебирать различные комбинации индексов главных компонент и соответствующие им подпространства  $MTX$  и  $MTy-$  Для каждой пары  $(7; , Ty)$  индексов главных компонент рядов  $X$  и  $y$  соответственно будем находить среднее расстояние между  $k$ : ближайшими соседями для ряда  $X$  и между ближайшими соседями для ряда  $y$ . Введем меру близости векторов

$_ dist(X, 7;) _ _ 1 k$   
 $S(Xay77; c77; r> — m: d15t(Xa7; c> — _ Z „ XI _ Xbly (2)$

Тогда задача поиска подпространств № $7;$  и  $MTy$  эквивалентна поиску номеров главных компо—нент  $(7; , 73,)$  и имеет вид  
 $(“Г...Ty) ””?-3,2% $( x,y,Tx,7y>, (3)$

ПЦ % min,  
ПИ % min 66  
4 Алгоритм многомерной гусеницы (MSSA-L)

$... „ T$   
В данной работе в качестве прогностической модели  $f 3$ —мерного ряда  $X : [k..., \dots, X(“”)]$  66 исполь—зуется алгоритм многомерной гусеницы (MSSA—L). Функция  $f$  имеет вид:

$... \dots \$\$“$   
 $ШЕ—b—2$   
 $2 2$   
 $за!... .. mi ; >$   
 $f<w7xt7Xt—17””7Xt—L+2> : _ p$   
 $(s) (s)$   
 $33t_L+2 ... slit$

вектор коэффициентов  $p$  определяется алгоритмом многомерной гусеницы MSSA—L. Алгоритм MS SA—L подробнее описан в следующем разделе.  
Алгоритм MMSA—L является обобщением на многомерный случай алгоритма гусеницы (SSA).

Задача алгоритма MSSA—L состоит в представлении временного ряда в виде суд/путы интерпретируемых компонент. Это осуществляется в четыре шага: запись ряда в виде траекторной матрицы, сингулярное разложение этой матрицы, группировка компонент, полученных при сингулярном разложении, в интерпретируемые компоненты 1/1 восстановление временного ряда по каждой из интерпретируемых компонент.

По ряду (1) построим матрицу Ганкеля  $H \in \mathbb{R}^{L \times K}$ ,  $K : N \rightarrow L + 1$ :

$H : [H_{17}H_{27} \dots H_{N3}]^T$

где  $L$  \* ширина окна,  $H_i \in \mathbb{R}^{L \times K}$  \* матрица Ганкеля для ряда  $x^{(i)}$ ,

$mg \gg mg \dots 15$ ;

(i) (i) (i)

$H_{i+1} \dots 331 \ 332 \ \dots N_{L+1}$

(i) (i) (i)

$33L_1$  жд  $\dots jN_{\mathbb{N}_1}$

По матрице Ганкеля  $H$  можно восстановить временной ряд  $X$ . Метод многомерной гусеницы строит приближение  $H$  матрицы  $H$  меньшего ранга с помощью сингулярного разложения этой матрицы 1/1 восстанавливает ряд по матрице  $H$ . Сингулярное разложение матрицы  $H$  имеет вид

$$d$$

$H : UAV : E \text{ inuiva}$

$i=1$

где  $A_1, \dots, A_C; > O_i$  сингулярные числа матрицы  $H$ ,  $\sigma_i$  1/1  $V_i$  \* столбцы матриц  $U \ H \ V$ . Тогда

наилучшее приближение матрицы  $H$  матрицей ранга  $T < d$  имеет вид :

7...

$H: E \text{ Aiuivg}$ —

$i=1$

По матрице  $H$  восстанавливается временной ряд  $X$  путем усреднения элементов, стоящих на антидиагоналях.

АЛГОРИТМ 1V1НОГО1V16РНОЙ гусеницы T&KЖ6 ПОЗВОЛЯЕТ ПОСТРОИТЬ ПРОГНОЗ временного ряда  $B \in \mathbb{N}_{10}$ —мент  $N \ 110$  ( $L \rightarrow 1$ ) предыдущим значениям ряда. Алгоритм находит такой вектор коэффициен—

тов  $p \in \mathbb{R}^{(L-1)}$ , что значения ряда  $X$  в момент  $N$ :

(1) (1)

$37N \rightarrow L-1 \ \text{дБм} \rightarrow 1$

(2) (2)

$LEN\_L\_1 \dots CEN\_1$

$XN \rightarrow p : Y - p$  (4)

A31", 335511

Заметим, что коэффициенты  $p$  оказываются общими для всех компонент ряда  $X$ .

Для каждого  $z' \in [1,74]$  обозначим  $f_1, 116p_{BB16}$  ( $L \rightarrow 1$ ) компонент столбца и,  $7T, 7$  последнюю

$T \ 2$

компоненту столбца  $u$ , и  $i : 22.21 \ 7TZ$ . Тогда вектор коэффициентов  $p$  вычисляется по формуле:

$1 \ 7'' \dots$

$p : \_1 \rightarrow y_2 \ \text{gnu}, (5)$

ВамGTI/NI, ЧТО ДЛЯ ОДНОЪЛЭРНОГО временного РЯД& СПР&В&ДЛИВЪ1 все приведенные СООТНОШЕНИЯ приз=1.

5 Метод проекций на патентные структуры

Пусть поставлена задача построения прогноза временного ряда  $X$  на несколько моментов вперед.

Пусть  $X \in \mathbb{R}^{n \times p}$  траекторная матрица ряда  $x$ ,  $Y \in \mathbb{N}^m \dots T \ 7$  значения ряда в последующие моменты

времени. Предполагается, что между строками матриц  $X$  и  $Y$  существует линейная зависимость:

$Y_i = X_i \cdot \alpha + e, X, 6R'', Y, E\mathbb{X} \ i=1, \dots, m,$

где  $\alpha$  \* матрица параметров модели, а  $\gamma$  вектор ошибок прогноза.

Ошибка прогноза вычисляется следующим образом:

$\$ < 9, X, Y > : NY \rightarrow X \cdot \alpha \ N\$ = \text{Хива} \rightarrow Xz \sim - \cdot @113$

Для нахождения параметров модели  $\alpha$  используется метод проекций на латентные структуры

PLS. Алгоритм PLS находит в латентном пространстве матрицу  $T \in \mathbb{R}^m$ », наилучшим образом опи—

сывающую матрицы  $X$  и  $Y$ . Матрицы  $X$  и  $Y$  проецируются в латентное пространство следующим

образом:



$$X = T \cdot P^T + F = Z \cdot t_k \cdot p_g + F$$

$$Y = Z \cdot T \cdot Q^T + E = Z \cdot t_k \cdot q_i + E$$

где  $T$  — матрица совместного описания объектов и ответов в латентном пространстве, причём столбцы матрицы  $T$  ортогональны;  $P, Q$  — матрицы перехода из латентного пространства в исходные пространства;  $E, F$  — матрицы невязок. Алгоритм PLS находит матрицы  $T, P, Q$ , а также такую матрицу  $W$ , что параметры модели можно вычислить по формуле

$$b = W(P^T W)^{-1} Q^T$$

### 6 Вычислительный эксперимент

#### 6.1 Сгенерированные данные

Эксперимент проводился на двух сгенерированных рядах  $X$  и  $Y$ :

$$X = 51 \cdot \sin(t) + 251 \cdot \sin(5t) + 0.3557 \cdot t$$

$$Y = 8111(225 + 5) + 0.36 \cdot t; \quad t \in [0, 25]$$

где  $t$  — время,  $\sin$  — синус.

Строим матрицу Ганкеля  $H_X$  по ряду  $X$ , взяв ширину окна  $L = 250$ . Для некоторого момента времени  $t^*$  рассмотрим вектор  $X_{t^*}$ , равный  $t^*$ -й строке матрицы  $H_X$ . Выберем  $k$  и найдем среди строк матрицы  $H_X$   $k$  ближайших (в смысле евклидовой нормы) соседей вектора  $X_{t^*}$ . Обозначим индексы найденных векторов  $t_1, \dots, t_k$ , а сами найденные векторы  $X_{t_1}, \dots, X_{t_k}$ . На рисунке изображен ряд  $X$  и  $k = 25$  ближайших соседей для момента  $t^* = 15$ . Моменты времени  $t_1, \dots, t_k$  выделены красным, момент  $t^*$  — черным.

0 500 1000 1500 2000 2500

Рис. 1: Ближайшие соседи точки  $X_{t^*}$

Строим матрицу Ганкеля  $H_Y$  по ряду  $Y$ . Обозначим  $i$ -ю строку  $H_Y$  через  $y_i$ . Тогда по найденным индексам  $t_1, \dots, t_k$  можно отобрать соответствующие  $y_{t_1}, \dots, y_{t_k}$ . Если ряд  $X$  зависит от ряда  $Y$  то вектора  $y_{t_1}, \dots, y_{t_k}$ , как  $1/1$  хи,  $t_1, \dots, t_k$ , будут находиться рядом в траекторном пространстве. Изобразим  $H_X$ ,  $H_Y$ :  $1/1$  уд,  $t_1, \dots, t_k$  на фазовых траекториях.

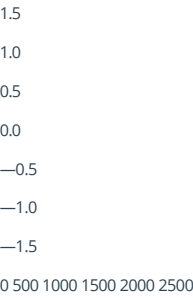
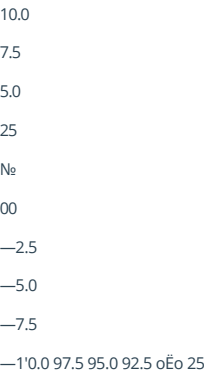


Рис. 3: Ближайшие соседи точки  $y_{t^*}$

25 neighbors, X 25 neighbors, Y



Ш

Рис. 2: Точки на фазовых траекториях рядов  $X$  и  $Y$ , соответствующие 25-ти ближайшим соседям  $X_{t^*}$ . Видно, что точки обоих фазовых траекториях расположены близко друг другу. Значит, ряд  $X$  зависит от ряда  $Y$ .

Аналогично для некоторого момента  $t^*$  найдем 25 ближайших соседей вектора  $X_{t^*}$ . Обозначим их  $t_1, \dots, t_k$ .

$t_1: t_k$

На рис. 3 изображен ряд  $X$  и  $k = 25$  ближайших соседей вектора  $y_{20}$ .  
Изобразим  $y_t^*$ ,  $y_t$ ,  $\dots, y_{t+k}$  и соответствующие ИМ  $X_N$ ,  $X_{751}, \dots, X_{th}$  на фазовых траекториях (рис. 4).

25 neighbors, Y 25 neighbors, X  
100 -

Рис. 4: Точки на фазовых траекториях рядов  $X$  и  $y$ , соответствующие 25—ти ближайшим соседям  $y_{20}$   
Видим, что точки  $X_t^*$ ,  $X_{151}, \dots, X_{15}$  расположены на фазовых траекториях близко друг к другу.  
При этом они распадаются на четыре плотные группы. Это связано с тем, что 67 период ряда  $X$  в четыре раза меньше периода ряда  $X$ .

6.2 Эксперимент на данных потребления электроэнергии 1/1 температуры  
В эксперименте исследуются ряд объема потребления электроэнергии  $X$  и ряд значений температу—  
ры  $X$  в течение года. Так как эти ряды не являются стационарными, их необходимо продифферен—  
цировать и отнормировать перед тем, как исследовать зависимости между ними. Ряд температуры  
будем приводить к стационарной форме следующим образом. Рассмотрим ряд длины светового дня  
в течение года 2. С помощью вычисления кросс—корреляционной функции  $*u_x(h)$  рядов  $X$  и 2. Опре—  
делим, насколько ряд  $Z$  опережает ряд  $X$ . То есть найдем такое  $h^*$ , что  $X(t + h^*) = 205$ ). Вычтем из  
ряда  $X$  ряд  $Z$  с учетом сдвига  $h^*$ . Полученный ряд  $X(t) = X(t) - z(t - N_0)$  будет стационарным рядом  
температуры.

Исходные ряды потребления электроэнергии  $x$ , температуры  $X$  и длины светового дня 2 изобра—  
жены на рис. 5.

— Electricity cons.  
— Temperature  
— Day length  
4000 5000 8000  
Time, hours

Рис. 5: Нормированные ряды потребления электроэнергии, температуры и длины светового дня  
Построим кросс—корреляционную диаграмму рядов  $x$  и  $z$ .

0.75  
0.50  
0.25  
0.00  
Q  
N  
>>  
>  
—0.25  
—0.50  
—0.75  
—1.00

Рис. 6: Кросс—корреляционная диаграмма для ряда температуры  $x$  и длины светового дня  $z$   
Максимум модуля кросс—корреляции  $\gamma_{yz}(h)$  достигается при  $h = 560$ . Значит,  
 $z(t) = X(t - | - 560)$ .  
И новый стационарный ряд температуры имеет вид  
 $x^*(i) : X(t) - z(t - 560)$ .

Далее, для удобства, полученный ряд температуры  $x^*$  будем обозначать  $x$ . Продифференцирован—  
ные и нормированные ряды потребления электроэнергии и температуры изображены на рис. 7.  
Исследуем зависимость ряда температуры  $y$  от ряда потребления электроэнергии  $x$ . Делаем это  
аналогично ЭКСПЭРИМЕНТУ на ИСКУССТВЕННЫХ данных. Выбираем ШИРИНУ ОКНА  $L$  и  $N$  КОТОРЫЙ момент

— Electricity cons.  
— Temperature  
0 50 100 150 200 250 300 350 400  
Time, hours

Рис. 7: Продифференцированные И нормированные ряды потребления электроэнергии и темпера—  
туры  
времени 15\*. Находим  $k$  ближайших соседей векторов  $x_{t^*}^{1/1} y_{t^*}$ , и их расположение в траекторном

пространстве.

Возьмем  $L = 170$ , что соответствует периоду в семь дней. Возьмем  $15^* = 400$ . На рис. 8 красным показаны ближайшие соседи вектора  $x_d$

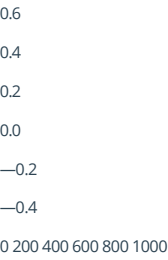


Рис. 8: Ближайшие соседи вектора  $x_d$ ... ширина окна  $L = 170$

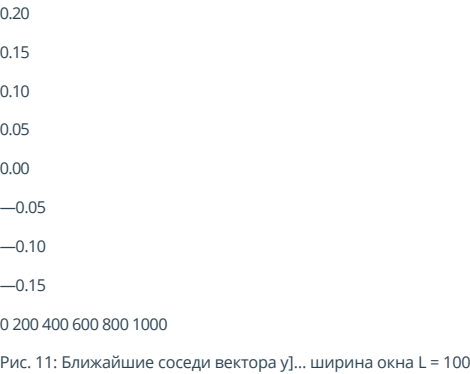
20 neighbors,  $X$  20 neighbors,  $Y$

Рис. 9: Вектора  $X_{t^*}$ ,  $X_{1517}$ , ...,  $x_d$  (ближайшие соседи вектора  $x_d$ ) И соответствующие вектора  $y_d$ ,  $y_{t^*}$ , ...,  $y_{1517}$  на фазовых диаграммах с периодом 12 часов

20 neighbors,  $X$  20 neighbors,  $Y$

Рис. 10: Вектора  $X_{t^*}$ ,  $X_{t^*+1}$ , ...,  $x_d$  (ближайшие соседи вектора  $x_d$ ) и соответствующие вектора  $y_d$ ,  $y_{t^*+1}$ , ...,  $y_{t^*+1517}$  на фазовых диаграммах с периодом 24 часа

Исследуем зависимость ряда  $x$  от ряда  $y$ . Возьмем  $15^* = 400$ ,  $L = 170$ . На рис. 11 красным показаны ближайшие соседи вектора  $y_d$



20 neighbors,  $Y$  20 neighbors,  $X$

0.3  
0.2  
0.1  
0.0  
 $N_e$   
70.1  
70.2  
70.3  
43.3 —0.2 43.1 0.0 0.11 0:2 0:3 —e.0 —i.5 —i.0 43.5 0.10 0:5 1:0 1:5 2:0  
.711 .711

Рис. 12: Вектора  $y_{t^*}$ , ...,  $y_{t^*+1517}$  (ближайшие соседи вектора  $y_d$ ) И соответствующие вектора  $x_{t^*}$ ,  $x_{t^*+1}$ , ...,  $x_d$  на фазовых диаграммах с периодом 12 часов

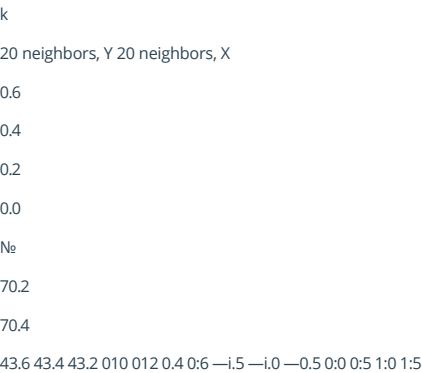


Рис. 13: Вектора  $y_{t^*}$ , ...,  $y_{t^*+1517}$  (ближайшие соседи вектора  $y_d$ ) и соответствующие вектора  $x_{t^*}$ ,  $x_{t^*+1}$ , ...,  $x_d$  на фазовых диаграммах с периодом 24 часа

6.3 Перебор подпространств

Переберем траекторные подпространства рядов  $X$  и  $y$  размера не больше пяти. Для этого будет перебирать пары множеств индексов главных компонент  $(7,; Tu)$ . Для каждой пары  $(7,; Tu)$  найдем  $S(X, y, 7,; Tu)$ , задающееся (2)

Рис. 14: Отношение расстояния между ближайшими соседями ряда  $X$  к расстоянию между соседями ряда  $y$ . Ближайшие соседи определяются по ряду  $X$

Рис. 15: Отношение расстояния между ближайшими соседями ряда  $y$  к расстоянию между соседями ряда  $X$ . Ближайшие соседи определяются по ряду  $X$

6.4 Построение прогноза

Построим прогноз ряда потребления электроэнергии  $X$ , используя только его собственную историю, и сравним с прогнозом, строящимся с использованием ряда температуры  $y$ . Рассмотрим также прогнозы, строящиеся при помощи только первых главных компонент ряда  $y$ . На рис. 16 представлены графики зависимости среднеквадратичной ошибки прогноза от ширины окна  $L$ .

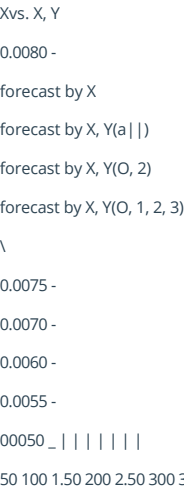
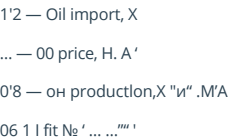


Рис. 16: Среднеквадратичная ошибка прогноза в зависимости от ширину окна

6.5 Эксперимент на данных об объеме перевозок нефти

В эксперименте проверяются связи между тремя временными рядами: объем грузоперевозок нефти  $x$ , цена на нефть  $h$  и объем добычи нефти  $q$ . Временные ряды заданы в период 21 год по месяцам. Исходные и продифференцированные ряды представлены на рис.17 и рис.18 соответственно.



W W "№0... "...

' V

0.2 ""

y" M' - "H".... "" <

0.0 "WWW' ''

0 50 100 150 200 250

Time, months

Рис. 17: Исследуемые временные ряды 66

2'0\_ — Oil import, X

— Oil price, H

1.5— — Oil production, Q

1.0

0.5

0.0

—{l5

—1.0

MEMEMMMMMWMMW,

III-IIIIIIWIIHII

MWMMMMMMM

III-III

NeNe | гNe ‘

нищим- |\_-

NeNeNeNeNeNe

и...

\_

пщNe

NeNeNeNeNeNeNeNeNeNe |

III-II

0 24

48 72 96

120 144

Time, months

168

IIIIIIII

192 216 240 264

Рис. 18: Исследуемые временные ряды, приведенные К стационарному виду

Исследуем связи рядов В парах (x, h) И (x, q). Аналогично предыдущим экспериментам, для проверки наличия связи между рядами смотрим, как точки, близкие в траекторном пространстве одного ряда, отображаются в траекторное подпространство другого ряда.

1.0-

0.5-

94

—0.5-

—1.0-

0.75-

0.50-

0.25-

92

—0.25-

—0.50-

—0.75-

—1.00

15 neighbors, X

0.0-

15 neighbors, H

0.75-

0.50-

0.25-

0.00-

—0.25-

—0.50-

—0.75-

0.00-

00 I . I I I I I I I

—0.8—0.6—0.4—0.2 0.0 0.2 0.4 0.6 0.8

Ш

15 neighbors, X

—0.8—0.6—0.4—0.2 0.0 0.12 0.14 0.6 0.18

Ш

Рис. 19: Связи между рядами x и h 66

15 neighbors, X 15 neighbors, Q

—1.0 —0.5 0.0 0.5 1.0 —1.5 —1.0 —0.5 0.0 0.5 1.0 1.5

уз 3J1

15 neighbors, Q 15 neighbors, X

—'1.5 —1.0 —0.5 0.0 0.5 1.0 1.5

зj1

Рис. 20: Связи между рядами X И q

Сравним ошибку прогноза ряда X С И без использования ряда h. Когда ряд h включен в прогно—

стическую модель рассмотрим два случая: для прогноза у ряда h, используются все компоненты,

используются только первые компоненты. Аналогично построим прогноз X С использованием q, h

С использованием X 1/1 q С использованием X. Результаты эксперимента представлены на рис. 22.

XV5.X,h HV5. H,X

| | | |

1...475 — forecast by X forecast by H

forecast by X, H(all) forecast by H, X(a| |)

forecast by X, H(1, 2) forecast by H, X(3, 4)

0,050 \*\*

0,0450 —

0,0425

0,045

0,0400

% 0,040

Ē

MSE

0.0375

0,035

0,0350

0.0325 0,030

0,0300

40 50 60 70 80 90 100 40 50 60 70 80 90 100

Рис. 21: Зависимость ошибки прогноза от ширины окна L Для пары рядов X 1/1 h. 66

Xvs.X,Q QVS- QX

0.010 '

forecast by X

forecast by X, @@") 66

forecast by X, Q(1y 2)

forecast by Q

forecast by Q, X(a| |)

forecast by Q, X(3, 4)

0.044 '

0.014 \*

0.042 '

0.012 -

0040\*

MSE

MSE

0.010-

0.0% -

0.1le-

шик; \_ WM

40 БН (... 70 >le 90 ШП 40 БН (... 70 >le 90 ШП

0.036 '

1...34 \*

Рис. 22: Зависимость ошибки прогноза от ширины окна L для пары рядов X и q.

7 Заключение

В работе решалась задача обнаружения связи между временными рядами, а также между их компо—

нентами. Связи между временными рядами устанавливались с помощью метода сходящегося пере—

крестного среднего (ССМ). Вывод о наличии связи между парой рядов сравнивался с результатом

теста Гренджера, проведенном на этих же рядах.

Эксперимент проводился на трех наборах данных: искусственные данные, данные потребления электроэнергии, данные РЖД объема грузоперевозок нефти. Для каждого набора данных сделаны выводы о наличии связей между исследуемыми рядами. На данных РЖД видно, что метод ССМ более чувствителен к наличию связи между рядами.

Работа по обнаружению связи между рядами с помощью теста Гренджера "Анализ зависимостей между показателями при прогнозировании объема грузоперевозок" подана в журнале "Системы и средства информатики".

Список литературы

[1] Clive WJ Granger. Investigating causal relations by econometric models and cross—spectral methods. **5**

Econometrica: Journal of the Econometric Society, pages 4247438, 1969.

[2] Adam B Barrett, Lionel Barnett, and Anil K Seth. Multivariate granger causality and generalized

variance. Physical Review **9** E, 81(4):041907, 2010.

[3] George Sugihara, Robert May, Hao Ye, Chih—hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan

Munch. Detecting causality in complex ecosystems. science, **5** page 1227079, 2012.

[4] George Sugihara and Robert M May. Nonlinear forecasting as a way of distinguishing chaos from

measurement error in time series. **59** Nature, 344(6268):734, 1990.

[5] Craig Hiemstra and Jonathan D Jones. Testing for linear and nonlinear granger causality in the

stock price—volume relation. The Journal of Finance **31**, 49(5):163971664, 1994.

[8]

[9]

[10]

[11]

[12]

[13]

[14]

[15]

[15]

[17]

[18]

[19]

Robert Hoffmann, Chew—Ging Lee, Bala Ramasamy, **18** and Matthew Yeung. Fdi and pollution: a granger causality test using panel data. Journal of international development **18**, 17(3):3117317, 2005.

Halbert White and Xun Lu. Granger causality and dynamic structural systems. Journal of Financial Econometrics, 8(2):1937243, 2010.

AM Katrutsa and VV Strijov. Stress test procedure for feature selection algorithms. Chemometrics and Intelligent Laboratory Systems, 142:1727183, 2015.

Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. ACM Computing Surveys (CSUR), 50(6):94, 2017.

Paul Geladi. Notes on the history and nature of partial least squares (pls) modelling. Journal of Chemometrics, 2(4):2317246, 1988.

Agnar Hoskuldsson. Pls regression methods. Journal of chemometrics, 2(3):2117228, 1988.

N Golyandina and D Stepanov. Ssa—based approaches to analysis and forecast of multidimensional time series. In proceedings of the 5th St. **1** Petersburg workshop on simulation, **2** volume 293, page 298, 2005.

Nina Golyandina, Vladimir Nekrutkin, and Anatoly A **6** Zhigljavsky. Analysis of time series structure: Ssa and related techniques (chapman & hall crc monographs on statistics & applied probability). 2001. **22**

Nina Golyandina and Anatoly Zhigljavsky. Singular Spectrum Analysis for time series. Springer **1** Science & Business Media, 2013.

James B Elsner and Anastasios A Tsonis. Singular spectrum analysis: a new tool in time series analysis. **16** Springer Science & Business Media, 2013.

Theodore Alexandrov. A method of trend extraction using singular spectrum analysis. arXiv preprint arXiv:0804.3367, 2008.

Myles R Allen and Leonard A Smith. Monte carlo ssa: Detecting irregular oscillations in the presence

of colored noise. *Journal of climate* **2**, 9(12):337373404, 1996.

Hossein Hassani, Saeed Heravi, and Anatoly Zhigljavsky. Forecasting uk industrial production With multivariate singular spectrum analysis. *Journal of Forecasting* **1**, 32(5):3957408, 2013.

CAF Marques, JA Ferreira, A Rocha, JM Castanheira, P Melo—Goncalves, N Vaz, **3** and JM Dias.

Singular spectrum analysis and forecasting of hydrological time series. *Physics and Chemistry of the Earth*, **3** Parts A/B/C, 31(18):117271179, 2006.