

Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»

На правах рукописи

Швыдун Сергей Владимирович

Суперпозиция в задачах анализа данных

РЕЗЮМЕ ДИССЕРТАЦИИ
на соискание ученой степени кандидата наук
по прикладной математике

Научный руководитель:
доктор технических наук,
старший научный сотрудник
Алескеров Фуад Тагиевич

Москва – 2020

Постановка проблемы

В последнее время большую актуальность приобретают задачи принятия индивидуальных, коллективных и многокритериальных решений, связанные с анализом больших объемов данных. Вследствие этого одной из важнейших характеристик процедур принятия решений становится их вычислительная сложность. К сожалению, большинство известных процедур имеют высокую сложность (не ниже квадратичной). Эта особенность приводит к необходимости применения приближенных процедур, имеющих низкую вычислительную сложность и значительно снижающих качество результатов.

Одним из способов решения проблемы высокой вычислительной сложности является применение принципа суперпозиции, который заключается в последовательном применении (композиции) различных процедур таким образом, что результаты применения одной процедуры подаются на вход другой процедуре выбора. Основным преимуществом моделей, основанных на принципе суперпозиции, является их низкая вычислительная сложность, так как, применяя на первых этапах процедуры с низкой вычислительной сложностью, можно существенно сузить исходный набор рассматриваемых вариантов. Модели, основанные на принципе суперпозиции, зачастую также имеют высокую степень интерпретируемости, поскольку представляют собой комбинацию простых методов вместо одного сложного. Тем не менее, поскольку операция суперпозиции не удовлетворяет свойству коммутативности, то есть полученные результаты могут существенно отличаться при изменении последовательности применения процедур принятия решений, одной из важнейших задач является исследование свойств различных моделей суперпозиции.

Степень разработанности проблемы

В 1900 г. Д. Гильбертом были сформулированы 23 наиболее важные проблемы, которые являлись наиболее значимыми в то время для

математики. Одной из них является задача о том, может ли функция нескольких переменных быть представлена посредством суперпозиции функций от меньшего числа аргументов (13-я проблема Гильберта). Самим Гильбертом была выдвинута гипотеза о невозможности данного представления и показано, что аналитические функции трех аргументов в общем случае не являются суперпозициями аналитических функций двух аргументов. В 1950-х гг. В.И. Арнольд и А.Н. Колмогоровым была показана неоднозначность решения 13-й проблемы Гильберта. В ряде своих работ А.Н. Колмогоров и В.И. Арнольд получили теоремы, опровергающие гипотезу Гильберта в классе непрерывных функций. В работе М.А. Айзермана и Ф.Т. Алескерова 1990 г. особое внимание было уделено различным функциям выбора, а также операциям (их композиции для создания новых функций) над ними. Авторами была исследована замкнутость областей, содержащих различные функции выбора, относительно операции суперпозиции. Помимо прочего, были рассмотрены различные типы двухступенчатых процедур выбора, а также определены условия, при которых данные процедуры сводятся к одноступенчатым. Ряд других работ М.А. Айзермана, А.В. Малишевского, В.И. Вольского, Ф.Т. Алескерова и Е. Чинара посвящен исследованию свойств некоторых двухступенчатых моделей выбора. Тем не менее, на данный момент рассмотрены лишь некоторые общие свойства двухступенчатых процедур выбора, а вопросы сложности данных процедур и их применения в задачах анализа больших данных практически нигде не рассматривались.

Цели и задачи исследования

Цель исследования. Целью данной работы является определение свойств различных моделей суперпозиции, а также построение на их основе моделей, имеющих высокую эффективность в различных прикладных исследованиях.

Задачи диссертационного исследования:

1. провести анализ зарубежных и отечественных работ, посвященных изучению различных моделей суперпозиции функций выбора;
2. определить список нормативных свойств, характеризующих рациональность функций выбора;
3. исследовать нормативные свойства существующих функций выбора, лежащих в основе построенных моделей суперпозиции;
4. исследовать нормативные свойства моделей суперпозиции, а также их вычислительную сложность;
5. построить эффективную модель поиска релевантных страниц, основанную на суперпозиции функций выбора.
6. построить эффективную модель предсказания возникновения торнадо, основанную на суперпозиции функций выбора.
7. построить эффективную модель распределения спорных зон в Арктическом регионе, основанную на суперпозиции функций выбора.

Актуальность исследования

Исследование свойств различных процедур выбора и их суперпозиций имеет огромное значение по ряду причин. Поскольку данные свойства позволяют определить, насколько устойчив и постоянен конечный набор наилучших вариантов, результаты анализа дают более детальное понимание основных особенностей, преимуществ и недостатков от применения тех или иных процедур выбора.

Изучение этих свойств важно также при анализе больших данных. В частности, информация о свойствах используемой многокритериальной процедуры позволяет понять, можно ли производить выбор не по всему множеству вариантов, а по его подмножествам, необходимо ли включать в анализ информацию о «наихудших» вариантах, а также в каких случаях

необходимо осуществлять перерасчет результатов, если значение для некоторых вариантов изменилось по одному из критериев. Немаловажную роль играет и изучение вычислительной сложности различных моделей суперпозиции с целью определения того, какие из них могут быть использованы при анализе больших объемов данных.

Наконец, на данный момент существует ряд областей, в которых задача выбора вариантов не может быть решена с достаточной точностью за приемлемое время в силу высокой вычислительной сложности существующих процедур. Эта особенность приводит к необходимости разработки новых методов принятия решений, в том числе основанных на принципе суперпозиции.

Личный вклад автора в разработку проблемы

Автором сформулированы и доказаны Теоремы 1-2 о свойствах известных процедур многокритериального выбора, а также построенных на их основе двухступенчатых моделей. Оценка вычислительной сложности процедур многокритериального выбора, приводимых в данном исследовании, получены автором лично.

Работа по предсказанию возникновения торнадо выполнена в соавторстве со студентами Н.Н. Байбородовым и С.С. Деминым, к.т.н. В.И. Якубой, д.т.н. Ф.Т. Алескеровым, а также профессорами Университета Оклахомы (США) М. Ричманом и Т. Трафалисом, в рамках которой автор занимался построением моделей, основанных на суперпозиции надпороговых процедур выбора, а также оценкой их эффективности.

Работа по предсказанию релевантности страниц выполнена в соавторстве со студентом Е.О. Митичкиным, к.т.н. В.И. Якубой, а также д.т.н. Ф.Т. Алескеровым, в рамках которой автор занимался задачей выявления ключевых показателей в исходных данных, построением моделей суперпозиции, их программной реализацией и оценкой эффективности по сравнению с другими известными моделями.

Работа по распределению спорных зон в Арктике выполнена в соавторстве с д.т.н. Алескеровым Ф.Т., в рамках которой автор занимался сбором и обработкой исходных данных, построением моделей распределения спорных зон, в том числе основанных на принципе суперпозиции, а также программной реализацией данных моделей.

Результаты диссертационного исследования докладывались на следующих научных семинарах, школах и конференциях:

1. 3-й международный семинар “Models of Influence and Network Theory”, г. Париж (Франция). Название доклада: "Over-lath superposition algorithm and its use to the ranking of options in the problem of search", 14.05.2012-15.05.2012;
2. Общественный семинар ИПУ РАН "Экспертные оценки и анализ данных", г. Москва (Россия). Название доклада: "Алгоритмы ранжирования, основанные на идее суперпозиции, и их применение к задаче информационного поиска", 28.11.2012;
3. Конференция ITQM 2013, г. Сучжоу (Китай). Название доклада: "Super-threshold Procedures and Their Application to the Search Problem", 16.05.2013-18.05.2013;
4. Научный семинар МЛАВР НИУ ВШЭ, г. Москва (Россия). Название доклада: "Некоторые процедуры выбора и их свойства", 10.06.2013;
5. 12-е Всероссийское совещание по проблемам управления, г. Москва (Россия). Название доклада: "Исследование нормативных свойств двухступенчатых процедур выбора", 19.06.2014;
6. Конференция IFORS 2014, г. Барселона (Испания). Название доклада: "Two-Stage Superposition Choice Procedures and their Properties", 13.07.2014-18.07.2014;
7. Общественный семинар ИПУ РАН "Экспертные оценки и анализ данных", г. Москва (Россия). Название доклада: "Двухступенчатые процедуры выбора и их свойства", 28.01.2015;

8. Конференция MCO 2015, г. Мец (Франция). Название доклада: "Properties and Complexity of Some Superposition Choice Procedures", 11.05.2015-13.05.2015;
9. Конференция EURO2015, г. Глазго (Великобритания). Название доклада: "Normative properties of the superposition of multi-criteria choice procedures", 12.07.2015-15.07.2015;
10. Конференция CYBCONF 2017, г. Эксетер (Великобритания). Название доклада: "A Mathematical Approach to Conflict Resolution in the Arctic Region", 21.06.2017-23.06.2017.
11. Конференция IFORS 2017, г. Квебек (Канада). Название доклада: "Conflict resolution models in the Arctic region", 17.07.2017-21.07.2017;
12. Конференция MLSO'2017, г. Москва (Россия). Название доклада: "Superposition Models of Conflict Resolution in the Arctic Region", 02.10.2017-04.10.2017;
13. UTFORSK Norwegian-Russian Workshop on Arctic Logistics, г. Молде (Норвегия). Название доклада: "Allocation of Disputable Zones in the Arctic", 10.10.2017-12.10.2017;
14. Семинар ЦЭМИ РАН "Математическая экономика", г. Москва (Россия). Название доклада: "Распределение спорных территорий в Арктическом регионе", 06.02.2018;
15. Конференция EURO2018, г. Валенсия (Испания). Название доклада: "On various solutions of areas allocation problem", 8.07.2018-11.07.2018;
16. Общественный семинар НИУ ВШЭ "Математические методы анализа решений в экономике, бизнесе и политике", г. Москва (Россия). Название доклада: "Модели суперпозиции в анализе данных", 17.10.2018;
17. Autumn school "Current trends in decision-making analysis", г. Москва (Россия). Название доклада: "Superposition Models in Data Analysis", 07.11.2018;

18.Школа SVF-8063 School of Society and Advanced Technology in the Arctic, г. Лонгьир (Норвегия). Название доклада: “Superposition (Composition) Models in Data Analysis”, 13.10.2019-19.10.2019.

Описание методологии исследования

Для теоретического исследования свойств суперпозиционных моделей принятия решений предполагается использование методов оптимизации, теории выбора, теоретико-игровых методов, теории вычислений и методов современной прикладной алгебры. Экспериментальный анализ подразумевает использование компьютерного, в том числе имитационного моделирования.

Теоретическая значимость работы заключается в

1. определении списка нормативных свойств, которым удовлетворяют существующие модели многокритериального выбора;
2. определении списка нормативных свойств, которым удовлетворяют двухступенчатые модели суперпозиции, основанные на известных процедурах многокритериального выбора;
3. исследовании вычислительной сложности рассматриваемых моделей суперпозиции;
4. разработке новых математических моделей поиска релевантных страниц, предсказания возникновения торнадо и распределения спорных территорий, основанных на суперпозиции различных функций выбора.

Практическая значимость диссертационного исследования

заключается в высокой эффективности предложенных моделей суперпозиции в задаче предсказания возникновения торнадо, задаче поиска, а также задаче распределения спорных территорий в Арктическом регионе. Кроме того, предложенные модели могут быть использованы в аналогичных задачах

прогнозирования в других областях знаний, а также при исследовании пространственных моделей взаимодействия между различными участниками в экономической и политической сферах.

Основные результаты исследования и положения, выносимые на защиту:

1. исследованы нормативные свойства 28 известных процедур многокритериального выбора и 617 двухступенчатых моделей суперпозиции и оценена их вычислительная сложность;
2. построены новые модели, основанные на суперпозиции функций выбора, показывающую высокую эффективность в задаче предсказания возникновения торнадо, задаче поиска, а также задаче распределения спорных территорий в Арктическом регионе.

Научная новизна. В рамках диссертационного исследования получены следующие новые научные результаты:

1. впервые исследованы свойства 28 известных правил многокритериального выбора и их суперпозиций на предмет удовлетворения существующим нормативным условиям;
2. впервые оценена вычислительная сложность исследуемых моделей суперпозиции и определен список моделей, которые могут быть применены при анализе большого числа вариантов.
3. построены новые модели суперпозиций функций выбора, показывающие высокую эффективность в задаче поиска, задаче предсказания возникновения торнадо, а также задаче распределения спорных территорий в Арктическом регионе.

Общие выводы и результаты исследования:

В рамках диссертационного исследования:

1. Исследованы нормативные свойства известных процедур выбора и их суперпозиций. Приводятся две теоремы о том, каким условиям

рациональности удовлетворяют изучаемые процедуры многокритериального выбора и их двухступенчатая суперпозиция.

Среди 28 процедур многокритериального выбора только правило надпорогового выбора, которое осуществляет выбор вариантов, имеющих значение по критерию выше некоторой константы, удовлетворяет всем рассматриваемым свойствам (кроме пороговой некомпенсируемости). Что касается остальных правил, то они не удовлетворяют большинству условий рациональности (кроме монотонности 1). Условие наследования выполняется только для правила простого большинства, а также правил, выбирающих недоминируемые варианты по мажоритарному отношению. Условие согласия выполняется лишь для некоторых правил, основанных на попарном сравнении вариантов. Только минимальное доминирующее множество и правило простого большинства удовлетворяют условию отбрасывания. Подробная информация о свойствах процедур многокритериального выбора приведена в Приложении Б диссертационной работы.

Среди двухступенчатых суперпозиций известных процедур многокритериального выбора лишь немногие правила удовлетворяют каким-либо условиям рациональности. Например, двухступенчатая процедура, на первом этапе которой применяется правило Фишберна, а на втором этапе правило простого большинства, удовлетворяет условиям наследования (Н), согласия (С), а также условию монотонности 1. В первую очередь, данная особенность может быть объяснена тем фактом, что большинство рассматриваемых функций выбора также не удовлетворяют данным нормативным условиям. Подробная информация о двухступенчатых процедурах, удовлетворяющих исследуемым нормативным условиям, приведена в Приложении С диссертационной работы.

Полученные результаты о свойствах рациональности существующих процедур многокритериального выбора и их суперпозиций имеют высокую значимость при анализе данных. Например, если процедура удовлетворяет условию монотонности 1, и в исходных данных произошло улучшение

оценок по критериям только для выбранных вариантов, то результаты пересчитывать не нужно. Другой пример - если процедура удовлетворяет условию C , то вычисление результатов можно производить на подмножествах вариантов, что позволяет снизить время работы процедуры.

Несмотря на то, что проведенный анализ показал, что большинство условий не выполняются для рассматриваемых процедур, был также выявлен ряд правил, удовлетворяющий некоторым интересным свойствам, что открывает дорогу для их активного использования в задачах анализа данных.

2. Оценена вычислительная сложность моделей суперпозиций, порожденными процедурами многокритериального выбора. Показано, что модели суперпозиции, имеющие вычислительную сложность не ниже квадратичной, не могут быть использованы на первом этапе суперпозиции при анализе больших объемов данных. Определен список процедур, сложность которых сильно зависит от общего числа вариантов, оставшихся в рассмотрении по после выполнения 1-го этапа. Выявлен ряд двухступенчатых моделей суперпозиции, которые могут применяться при наличии большого числа вариантов. Кроме того, проведено тестирование некоторых процедур на сгенерированных случайным образом данных и оценено время их выполнения.

3. Предложены три модели, основанные на принципе суперпозиции, которые использованы в задаче поиска, задаче предсказания возникновения торнадо, а также задаче распределения спорных территорий в Арктическом регионе.

В рамках задачи по предсказанию возникновения торнадо была предложена модель, основанная на суперпозиции надпороговых процедур, которая производит исключение наблюдений по 1 или 2 параметрам на каждой ступени. В результате проведенного тестирования модель суперпозиции показала более высокую эффективность по различным метрикам качества по сравнению с другими наиболее используемыми в

литературе методами (логистическая регрессия, метод опорных векторов, случайный лес и т.д.).

В рамках задачи поиска был разработан алгоритм, который позволяет определить релевантность документов. Алгоритм также основан на суперпозиции надпороговых процедур и отличается простотой, а также достаточно высокой точностью прогнозирования релевантности по сравнению с другими известными методами (деревья решений, метод опорных векторов).

Наконец, разработано несколько моделей распределения зон для разрешения конфликта интересов в регионе, которые могут быть использованы в переговорах по принятию решений и урегулированию спорных вопросов в данном регионе. Результаты применения моделей оказались практически идентичными (совпадение на 90%), но модель, основанная на суперпозиции, позволяет получить результаты за меньшее число итераций. Стоит отметить, что целью работы является разработка математических моделей, а не утверждение каких-либо сценариев распределения территорий. Ранний прогноз зон совместного интереса и дискуссии относительно их возможного распределения может облегчить процесс принятия решений в Арктическом регионе.

Таким образом, модели суперпозиции показали свою эффективность в различных задачах анализа данных.

Список опубликованных статей, где отражены основные научные результаты диссертации.

Публикации в международных журналах квартиля Q1, индексируемых в базах Web of Science и Scopus:

1. Aleskerov, F., Shvydun, S. Allocation of Disputable Zones in the Arctic Region. Group Decis Negot 28, 11–42 (2019).

Публикации в журналах/изданиях, индексируемых в базах Web of Science и Scopus:

2. Shvydun S., Aleskerov F., "A Mathematical Approach to Conflict Resolution in the Arctic Region", 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), Exeter, 2017, pp. 145-151.
3. Shvydun S. (2015) Properties and Complexity of Some Superposition Choice Procedures. In: Modelling, Computation and Optimization in Information Systems and Management Sciences. Advances in Intelligent Systems and Computing, vol 360. Springer, Cham. P. 475-486.
4. Aleskerov F., Mitichkin E., Shvydun S., Yakuba V. Super-threshold Procedures and Their Application to the Search Problem // Procedia Computer Science. 2013. No. 17. P. 1121-1124.

Прочие публикации:

5. Aleskerov, F., Demin, S. & Shvydun, S. Superposition of Choice Functions and Its Application to Tornado Prediction and Search Problems. SN COMPUT. SCI. 1, 68 (2020).
6. Shvydun S. V. Superposition Models of Conflict Resolution in the Arctic Region // В кн.: Материалы 10-й международной конференции "Управление развитием крупномасштабных систем (MLSD'2017)", М.: ИПУ РАН, 2017. P. 452-455.
7. Aleskerov F., Vaiborodov N., Demin S., Richman M., Shvydun S., Trafalis T., Yakuba V. Построение эффективной модели машинного обучения для предсказания торнадо. [Текст]: препринт WP7/2016/05. М.: Изд. дом Высшей школы экономики, 2016.
8. Shvydun S. V. Normative properties of multi-criteria choice procedures and their superpositions: I [Электронный ресурс]: препринт WP7/2015/07 (Часть 1), М.: Изд. дом Высшей школы экономики, 2015.
9. Shvydun S. V. Normative properties of multi-criteria choice procedures and their superpositions: II [Электронный ресурс]: препринт WP7/2015/07 (Часть 2), М.: Изд. дом Высшей школы экономики, 2015.

10. Швыдун С. В. Исследование нормативных свойств двухступенчатых процедур выбора // В кн.: XII Всероссийское совещание по проблемам управления ВСПУ-2014. М.: ИПУ РАН, 2014. С. 7977-7985.

Авторские свидетельства, патенты:

1. Fuad T. Aleskerov, Evgeny O. Mitichkin, Vyacheslav V. Chistyakov, Sergey V. Shvydun, Viacheslav I. Yakuba. USA Patent #US10275418 B2 «Method for selecting valid variants in search and recommendation systems», дата регистрации 30.04.2019.
2. Изобретение №2543315 «Способ отбора эффективных вариантов в поисковых и рекомендательных системах (варианты)», приоритет изобретения 22.03.2013, дата регистрации 27.01.2015 (соавторы: Алескеров Ф.Т., Митичкин Е.О., Чистяков В.В., Якуба В.И.).
3. Программа для ЭВМ №2013618228 «Выбор и ранжирование вариантов с использованием суперпозиции позиционных правил», дата государственной регистрации: 04.09.2013 (соавторы: Алескеров Ф.Т., Митичкин Е.О., Якуба В.И.).

Результаты исследования использовались в следующих грантах и научно-исследовательских проектах:

1. Грант РФФИ №12-01-00226 «Модели выбора, основанные на суперпозиции», руководитель: Алескеров Ф.Т., 2012-2014 гг.
2. Исследовательский проект «Математические модели принятия решений в социальных и экономических системах», руководитель проекта: Алескеров Ф.Т., ответственное подразделение: департамент математики факультета экономических наук НИУ ВШЭ, 2010 г.
3. Исследовательский проект «Конструирование экономических механизмов», руководитель проекта: Алескеров Ф.Т., ответственное подразделение: МЛАВР НИУ ВШЭ, 2011 г.

4. Исследовательский проект «Исследование новых методов и подходов в области математического моделирования и дизайна механизмов в социальной, экономической и политической сферах», руководитель: Алескеров Ф.Т., Маскин Э., ответственное подразделение: МЛАВР НИУ ВШЭ, 2013 г.
5. Исследовательский проект «Теоретическое и численное исследование современных математических моделей в социально-экономической, политической и финансовой сферах», руководители проекта: Алескеров Ф.Т., Маскин Э., ответственное подразделение: МЛАВР НИУ ВШЭ, 2014 г.
6. Исследовательский проект «Анализ данных и принятие решений в социально-экономических и политических системах», руководители проекта: Алескеров Ф.Т., Маскин Э., ответственное подразделение: МЛАВР НИУ ВШЭ, 2015 г.
7. Исследовательский проект «Анализ, выбор и принятие решений в социально-экономической, политической и финансовой сферах: новые модели, методы и алгоритмы», руководители проекта: Алескеров Ф.Т., Маскин Э., ответственное подразделение: МЛАВР НИУ ВШЭ, 2017 г.