



Корнилина Елена Дмитриевна

**Динамическая модель близости позиций индивидов
в замкнутых группах**

Специальность:

**05.13.18 – математическое моделирование, численные методы
и комплексы программ**

**Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук**

Москва – 2013

Работа выполнена в Институте прикладной математики им. М.В. Келдыша РАН

Научный руководитель: доктор физико-математических наук
ФГБУН Институт прикладной математики
им. М.В. Келдыша РАН, ведущий научный сотрудник
Петров Александр Пхоун Чжо

Официальные оппоненты: доктор физико-математических наук, профессор
ФГОБУ ВПО «Финансовый университет при Правительстве
Российской Федерации», заведующей кафедрой
«Прикладная математика»
Попов Виктор Юрьевич

кандидат технических наук, доцент
Московский институт электроники и математики
национального исследовательского университета «Высшая
школа экономики», доцент кафедры «Информационные
технологии и автоматизированные системы»
Клышинский Эдуард Станиславович

Ведущая организация: Федеральное государственное бюджетное учреждение науки
Институт системного анализа РАН

Защита диссертации состоится 23 мая 2013 г. в 14.00 часов на заседании диссертационного
совета Д 002.024.03 в Федеральном государственном бюджетном учреждении науки
Российской академии наук Институте прикладной математики им. М.В. Келдыша по адресу:
125047, г.Москва, Миусская пл., д.4.

С диссертацией можно ознакомиться в библиотеке ИПМ им. М.В. Келдыша РАН
Автореферат разослан 22 апреля 2013 г.

Учёный секретарь
диссертационного совета Д 002.024.03,
доктор физико-математических наук



Н.В. Змитренко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность диссертационной работы

Настоящая работа посвящена разработке динамической модели близости позиций участников замкнутых групп, высказывающих своё мнение по заданному набору вопросов.

Рассматривается группа с фиксированным числом участников (т.е. замкнутая), обсуждающих ряд вопросов таким образом, что мнение каждого становится сразу доступно всем остальным. Со временем позиции участников могут меняться под действием высказываний других членов группы. Между участниками имеются аффективные отношения (один участник может относиться к другому положительно или отрицательно), которые, в частности, влияют на изменение позиций индивидов по обсуждаемым вопросам. Эти отношения также могут меняться с течением времени.

Примером подобных взаимодействий могут служить обсуждения на различных форумах, в интернет-сообществах и т.п. В связи с существенным развитием популярности обмена мнениями в интернете, исследование взаимодействий данного типа, в том числе методами математического моделирования, становится особенно важным в настоящее время, что свидетельствует об актуальности настоящей работы.

Математические модели групповой динамики изучаются по крайней мере с 70-х годов двадцатого века, в них рассматриваются процессы изменения индивидуальных мнений под влиянием обсуждений в пределах замкнутой группы. Отметим несколько результатов моделирования установившихся (статических) групповых отношений. Согласно теории структурного баланса, предложенной Ф. Хайдером¹, входящий в группу индивид находится в состоянии когнитивного баланса, если для каждой триады, в которую он входит, его отношения с каждым из других членов этой триады, имеют одну из следующих форм: друг моего друга – мой друг (R1), враг моего врага – мой друг (R2), друг моего врага – мой враг (R3), враг моего друга – мой враг (R4). С точки зрения теории Хайдера, устойчивые триады – это те, в которых все участники близки между собой, либо двое близки друг другу и далеки от третьего. В теореме Хайдера-Картрайта-Харари^{2,3} показано, что в группе, в которой участники попарно обмениваются мнениями (положительными или

¹ Heider F. Attitudes and Cognitive Organization. - J. of Psych., 21, 1946, p. 107-112

² Cartwright D., Harary F. Structural Balance: A Generalization of Heider's Theory. - Psych. Rev., 63, 1956, p. 277-293

³ Kemeny J.G., Snell J.L. Mathematical Models in the Social Sciences. - New York: Blaisdell Publishing Co., 1962; reprinted by M.I.T. Press, Cambridge, Mass, 1972

отрицательными), стабильное состояние приводит к разбиению на две устойчивые подгруппы.

Сбалансированность по Дж. Дэвису^{4,5} является обобщением предыдущего понятия сбалансированности. Она основана на подходе, в рамках которого предлагается понимать триаду стабильной также и в том случае, когда все участники враждуют. Такой подход позволяет моделировать ситуации, в которых сформированы несколько антагонистических подгрупп, находящихся в стабильном состоянии.

Для модели, предлагаемой в настоящей работе, априори не было очевидно, какой тип сбалансированности группы в ней реализуется, более частный или более общий. В результате проведённых исследований выяснилось, что модель относится к более общему типу (как и приведённая выше модель Дэвиса), что позволяет реализовывать более широкий класс стабильных состояний.

Из наиболее близких динамических моделей к предлагаемой в работе, можно назвать модель Дж. Хантера⁶, В. Ковчегова⁷, П. Киллворта-Г. Бернарда⁸. Также можно отметить работу отечественных авторов Д.А. Губанова, Д.А. Новикова, А.Г. Чхартишвили⁹, в которой рассматриваются вопросы распространения информации в социальных сетях, что близко исследуемой тематике.

Главной отличительной чертой предлагаемого подхода от перечисленных работ является то, что рассматриваются не только попарные отношения. Мнение каждого индивида доступно всем остальным, в этом заключается массовость взаимодействий. В то же время, к каждому участнику группы остальные имеют индивидуальное отношение (положительное или отрицательное), поэтому имеет место попарно-групповой тип коммуникации, являющийся новым для группового моделирования.

Вторая часть работы посвящена получению данных для насыщения модели. Главные величины, на которые опирается модель – это отношения между участниками (субъективная дистанция, на которой один участник

⁴ *Davis J.A.* Clustering and Structural Balance in Graphs // *Hum. Relat.* 1967 V.20 p.181-188.

⁵ *Davis J.A., Leinhardt, S.* The Structure of Positive Interpersonal Relations in Small Groups // *Sociological theories in progress.* V.2 Boston, 1972, p.218-251.

⁶ *Hunter J.E.* Dynamic sociometry // *J. Math. Sociol.* 1978. V.6. p.87-138.

⁷ *Ковчегов В.Б.* Модель динамики групповых структур человеческих сообществ // *Социология* 4М. 1991. № 1. С. 75-98.

⁸ *Killworth P.D., Bernard H.R.* A Model Of Human Group Dynamics // *Social Science Research.* Vol. 5. 1976. P. 173-224

⁹ *Губанов Д.А., Новиков Д.А., Чхартишвили А.Г.* «Социальные сети: модели информационного влияния, управления и противоборства», 2010 – 228 стр.

воспринимает другого) и их позиции по каждому из обсуждаемых вопросов.

По вопросам измерения отношений между индивидами в группах можно назвать эмпирические работы по социометрии, например, статьи Дж. Морено¹⁰ и более поздние работы. Новые подходы к измерению отношений между индивидами в настоящей работе не предлагаются.

Для случая, когда участники обсуждают вопросы, имеющие отношение исключительно к политической тематике, имеется целый ряд методов по измерению близости позиций. Например, существует множество методов, основанных на анализе результатов голосований (см., например, работу Н. Благовещенского¹¹ и др.). Если говорить об анализе позиций, выраженных в политических текстах, то традиционные методы основываются на применении вариаций контент-анализа, требующего экспертной оценки. В частности, одна из наиболее известных методик анализа политических программ представлена в проекте “Manifesto”¹², её основным элементом является специального вида рубрикатор. Есть и другие методы, основанные на контент-анализе^{13,14,15}.

Естественным образом возникает задача разработки подхода к измерению близости позиций без привлечения эксперта, решение которой предлагает подход в настоящей работе, основанный на анализе политических текстов с помощью латентно-семантического анализа (ЛСА). Возможность применения данного подхода возникла в связи с предложенной гипотезой о том, что тексты, выражающие более близкие позиции (например, политические), будут синтагматически более близкими. Синтагматическая близость, в свою очередь, лежит в основе ЛСА¹⁶.

Цель работы – построение и исследование математической модели динамики близости позиций участников замкнутых групп, открыто высказывающих своё мнение по фиксированному набору вопросов.

¹⁰ *Moreno, J. L.* Sociometry, Experimental Method and the Science of Society. An Approach to a New Political Orientation. Beacon House, Beacon, New York – 1951.

¹¹ *Благовещенский Н.Ю.* Оценка политических позиций в Государственной Думе 3-го созыва по результатам голосований. М.: Фонд ИНДЕМ, 2004.

¹² *A.Volkens, O.Lacewell, P.Lehmann, S.Regel, H.Schultze, A.Werner* (2011): The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR), Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).

¹³ *Lowe W., Benoit K., Mikhaylov S., Laver M.* Scaling Policy Preferences From Coded Political Texts. Legislative Studies Quarterly 26(1, Feb): 123-155.

¹⁴ *Laver M., Benoit K., Garry J.* (2003). Extracting policy positions from political texts using words as data. American Political Science Review, 97(2), 311–331.

¹⁵ *Martin L.W., Vanberg G.*, A robust transformation procedure for interpreting political text, Political Analysis 16 (1), 93-100.

¹⁶ *Митрофанова О.А.* Семантические расстояния: проблемы и перспективы // XXXIV Международная филологическая конференция: Вып. 21. Прикладная и математическая лингвистика. СПб., 2005.

На защиту выносятся следующие результаты:

1. Разработана новая модель динамики близости позиций взаимодействующих индивидов в замкнутой группе, имеющая вид системы нелинейных обыкновенных дифференциальных уравнений для переменных, характеризующих отношения между индивидами, а также их позиции по обсуждаемым в группе вопросам.

2. Для построенной модели изучена зависимость решения от входных данных, доказана теорема об асимптотической устойчивости состояния согласия, т.е. состояния, в котором позиции всех индивидов по рассматриваемым темам одинаковы. Для базового случая модели доказана теорема о достаточных условиях существования неограниченных решений, соответствующих процессу расхождения позиций индивидов. Численным экспериментом показано существование решений, соответствующих формированию в группе нескольких подгрупп, в пределах каждой из которых достигается состояние согласия.

3. Разработан программный комплекс, реализующий предложенный эмпирический метод численного измерения близости политических позиций. Метод основан на предложенной в работе гипотезе об определении близости политических позиций через синтагматическое (т.е. проявляемое через контекст) расстояние между текстами, выражающими эти позиции. Проведена апробация программного комплекса на экспериментах по измерению близости позиций, представленных в предвыборных программах основных политических партий РФ в 2007 и 2011 гг.

Научная новизна результатов:

1. в предложенной модели новизна заключается в том, что с её помощью описываются попарно-групповые коммуникации и предмет моделирования – позиции участников, а не только отношения между ними;
2. показано, что в модели, описывающей группы, в которых участники обсуждают набор вопросов, а не только отношения друг к другу, возможно возникновение состояния согласия;
3. в основе реализованного программного комплекса лежит предложенный метод численного измерения близости политических позиций, который базируется на семантическом, а не контент-анализе;

Все представленные результаты являются новыми.

Практическая значимость результатов заключается прежде всего в том, что разработанная динамическая модель может быть использована при исследовании социально-политических процессов методами математического

моделирования.

В частности, модель может быть использована для анализа и прогнозирования изменения близости позиций членов политических сообществ, например, среди пользователей интернета.

Апробация работы. Результаты исследований докладывались на научных семинарах: расширенный семинар лаборатории динамических интеллектуальных систем ИСА РАН, 16 апреля 2013; научный семинар «Математическое моделирование» под рук. В.Ф. Тишкина и А.А. Кулешова, ИПМ им. М.В. Келдыша РАН, 21 июня 2012; научный семинар «Экспертные оценки и анализ данных» под рук. Ф.Т. Алескерова, ИПУ РАН, 23 мая 2012; научный семинар К.В. Рудакова, ВЦ РАН, 29 марта 2012; XIV Междисциплинарный ежегодный научный семинар «Математическое моделирование социальных процессов», МГУ ВМК, 22 декабря 2011; научный спецсеминар «Математические модели в гуманитарных науках», МГУ им. М.В. Ломоносова ВМК, 5 октября 2011; научный семинар «Эволюционные процессы в финансах и экономике» под рук. В.Ю. Попова и А.Б. Шаповала, Финансовый Университет, 18 мая 2011. А также на следующих конференциях: 6th International Conference on Performance Evaluation Methodologies and Tools, Cargese, France, 9-12 October 2012; международная научно-практическая конференция «Теория активных систем», 15 ноября 2011; XIV Всероссийская молодежная Конференция-школа с международным участием «Современные проблемы математического моделирования», Абрау-Дюрсо, 12-17 сентября 2011; XII и XIII Международная научная конференция по проблемам развития экономики и общества; VI Всероссийская научная конференция «Сорокинские чтения», Москва, 1-2 декабря 2010; VIII и IX International conference on Intellectualization of Information Processing; III International Conference on Mathematical Modeling of Social and Economical Dynamics, 23-25 June 2010; международная научная конференция «Ломоносов», 12-15 апреля 2010.

Работа выполнена при частичной финансовой поддержке РФФИ: гранты №№10-01-00332-а (И), 12-01-31461-мол_а (Р), 12-01-09369-моб_з (Р) и РГНФ: грант №12-03-00431-а (И).

Достоверность и обоснованность результатов. Достоверность научных результатов подтверждается математической строгостью при доказательстве утверждений и теорем. Результаты исследования построенной математической модели соответствуют известным теориям относительно отношений в группах в стационарном случае. Результаты, полученные с помощью разработанного программного комплекса соответствуют политологическим результатам.

Личный вклад. Все результаты, выносимые на защиту, получены

автором самостоятельно. Постановка задачи была выполнена совместно с научным руководителем.

Публикации. По результатам исследований подготовлено 14 публикаций, в том числе 3 публикации, включённые в Перечень ведущих рецензируемых научных журналов и изданий [1-3], 1 статья в журнале [12], 5 статей в сборниках трудов международных научных конференций [4, 5, 7, 8, 9], 3 статьи в сборниках докладов [11, 13, 14], 2 статьи в сборниках трудов всероссийской научной конференции и семинара [6, 10].

Структура и объем работы. Диссертация состоит из оглавления, введения, трех глав, заключения и списка литературы. Содержание работы изложено на 105 страницах.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы исследования, аргументируется научная значимость, характеризуется общая методологическая база работы, даётся постановка задачи, описывается состояние исследований в соответствующей области, приводится краткое изложение содержания работы.

В первой главе содержится постановка задачи моделирования изменения близости позиций, высказываемых участниками замкнутых сообществ; проводится формализация задачи; предлагаются поведенческие гипотезы, на основе которых строится математическая модель. Исследуются аналитически стационарные решения модели и их устойчивость. Проводится численное исследование устойчивых состояний модели в частных случаях.

В разделе 1.1 вводятся основные предположения о модели и поведенческие гипотезы, демонстрируются отличительные особенности модели.

Рассматривается замкнутая группа K индивидов, обсуждающих N вопросов. Обозначим через $x_{ij}(t) \geq 0$ субъективную дистанцию от i -го индивида до j -го; через $z_i^n(t) \in \mathbb{R}$ – позицию i -го индивида по n -му вопросу в момент времени t . Переменную $x_{ij}(t)$ следует понимать как некоторую совокупную величину, характеризующую субъективное расстояние, на котором воспринимает себя один участник по отношению к другому, при этом, вообще говоря, $x_{ij}(t) \neq x_{ji}(t)$. Будем полагать, что субъективная дистанция участника по отношению к самому себе имеет нулевое значение: $x_{ii}(t) = 0$, $i = 1, \dots, K$.

Позицию $z_i^n(t)$ участника можно понимать как положительное или отрицательное отношение к обсуждаемому вопросу.

Согласно условию модели, участники группы открыто высказывают

позиции по ряду вопросов, в связи с чем изменяются отношения между ними, а затем и их позиции. Будем основываться на следующих предположениях о связи субъективных дистанций и позиций:

1) позиции по n -му вопросу в момент времени t считаются достаточно близкими, чтобы способствовать улучшению отношений между участниками, когда $|z_i^n(t) - z_j^n(t)| < a$, где a – некоторый заданный положительный параметр;

2) если для большей части вопросов это верно, то субъективная дистанция уменьшается: $dx_{ij}(t) / dt < 0$;

3) позиции по рассматриваемым вопросам i -го участника притягиваются к позициям k -го участника, если они находятся внутри некоторой зоны восприятия его как единомышленника: $x_{ik} < R_i$, где $R_i > 0$;

4) влияния других членов на позицию каждого участника суммируются.

Данные предположения записаны в виде следующих уравнений модели:

(1)-(3)

$$\frac{dx_{ij}}{dt} = \gamma_i x_{ij}(t) \sum_{n=1}^N \alpha_i^n \left[\left(z_i^n(t) - z_j^n(t) \right)^2 - a^2 \right] \quad (1)$$

$$\frac{dz_i^n}{dt} = \sigma_i \sum_{j=1}^K \beta_{ij}^n (z_j^n(t) - z_i^n(t)) (R_i - x_{ij}(t)) \quad (2)$$

$$x_{ij}(0) = x0_{ij}, \quad z_i^n(0) = z0_i^n \quad (3)$$

$$i, j = 1, \dots, K, \quad i \neq j, \quad n = 1, \dots, N.$$

Параметр $\alpha_{in} \geq 0$ описывает субъективную важность n -го вопроса с точки зрения i -го индивида, $\sum_{n=1}^N \alpha_i^n = 1$, а параметр $\gamma_i \geq 0$ характеризует индивида с точки зрения того, насколько быстро или медленно он меняет свое отношение к другим членам группы.

Параметр $\sigma_i \geq 0$ описывает степень склонности индивида менять свою позицию по обсуждаемым вопросам, а параметр $R_i > 0$ – насколько близких участников он воспринимает как единомышленников. Воздействия на позицию i -го индивида по n -му вопросу со стороны всех других индивидов суммируются с весами $\beta_{ij}^n \geq 0$, нормированными так, что $\sum_{j=1}^K \beta_{ij}^n = 1$.

Таким образом, система уравнений модели динамики позиций имеет вид (1), (2) и представляет собой задачу Коши с начальными условиями (3). В общем случае, система имеет $K(K-1) + KN$ уравнений.

В разделе 1.2 доказывается теорема об устойчивости стационарного состояния, для некоторых частных случаев доказывается теорема о достаточных условиях неограниченного расхождения.

Введём дополнительное обозначение. Пусть $y_i^n(t) = z_i^n(t) - z_1^n(t)$, $i = 2, \dots, K, n = 1, \dots, N$. Тогда систему (1)-(2) можно переписать в следующем виде:

$$\frac{dx_{1j}}{dt} = \gamma_1 x_{1j} \sum_{n=1}^N \alpha_1^n [(y_j^n)^2 - a^2], j = 2, \dots, K \quad (4)$$

$$\frac{dx_{i1}}{dt} = \gamma_i x_{i1} \sum_{n=1}^N \alpha_i^n [(y_i^n)^2 - a^2], i = 2, \dots, K \quad (5)$$

$$\frac{dx_{ij}}{dt} = \gamma_i x_{ij} \sum_{n=1}^N \alpha_i^n [(y_i^n - y_j^n)^2 - a^2], i, j = 2, \dots, K \quad (6)$$

$$\begin{aligned} \frac{dy_i^n}{dt} = & -\sigma_i \beta_{ij}^n y_i^n (R_i - x_{i1}) \\ & + \sigma_i \sum_{j=2}^K \beta_{ij}^n (y_j^n - y_i^n) (R_i - x_{ij}) - \sigma_1 \sum_{j=1}^K \beta_{1j}^n (y_j^n) (R_i - x_{ij}), \end{aligned} \quad (7)$$

$i = 2, \dots, K, n = 1, \dots, N$

Определение 1. Назовём *состоянием согласия* системы (4)-(7) её стационарное решение, в котором $x_{ij} = x_{ji} = 0, i, j = 1, \dots, K, y_i^n = 0, i = 2, \dots, K, n = 1, \dots, N$.

Определение следует понимать следующим образом: в состоянии согласия каждый из индивидов считает другого единомышленником, и они занимают одну и ту же позицию по всем вопросам.

Теорема 1 (об устойчивости стационарного состояния). Пусть в системе (4)-(7) все коэффициенты $\alpha_i^n, \beta_{ij}^n, \gamma_i, \sigma_i, a, R_i$, где $i, j = 1, \dots, K, n = 1, \dots, N$, положительны. Тогда состояние согласия системы является асимптотически устойчивым.

Рассмотрим задачу

$$\left\{ \begin{array}{l} \frac{dx_1}{dt} = \gamma_1 x_1 [y^2 - a^2] \end{array} \right. \quad (8)$$

$$\left\{ \begin{array}{l} \frac{dx_2}{dt} = \gamma_2 x_2 [y^2 - a^2] \end{array} \right. \quad (9)$$

$$\left\{ \begin{array}{l} \frac{dy}{dt} = \sigma y [x_1 + x_2 - 2R] \end{array} \right. \quad (10)$$

$$\left\{ \begin{array}{l} x_1(0) = x_{10}, x_2(0) = x_{20}, y(0) = y_0 \end{array} \right. \quad (11)$$

где $y = z_1^1 - z_2^1$ и $x_{10} > R, x_{20} > R, y_0 > a$.

Теорема 2 (о разногласиях). Пусть коэффициенты $\gamma_1, \gamma_2, \sigma, a, R$ положительны и пусть $T > 0$ – число такое, что решение задачи Коши (8)-(11) существует на отрезке $0 \leq t \leq T$. Тогда неравенства

$$\begin{aligned} x_1(t) &> \alpha_1(t), \alpha_1(t) = x_1(0) \exp[\gamma_1(y^2(0) - a^2)t] \\ x_2(t) &> \alpha_2(t), \alpha_2(t) = x_2(0) \exp[\gamma_2(y^2(0) - a^2)t] \\ y(t) &> \alpha_3(t), \alpha_3(t) = y(0) \exp[\sigma(x_1(0) + x_2(0) - 2R)t] \end{aligned}$$

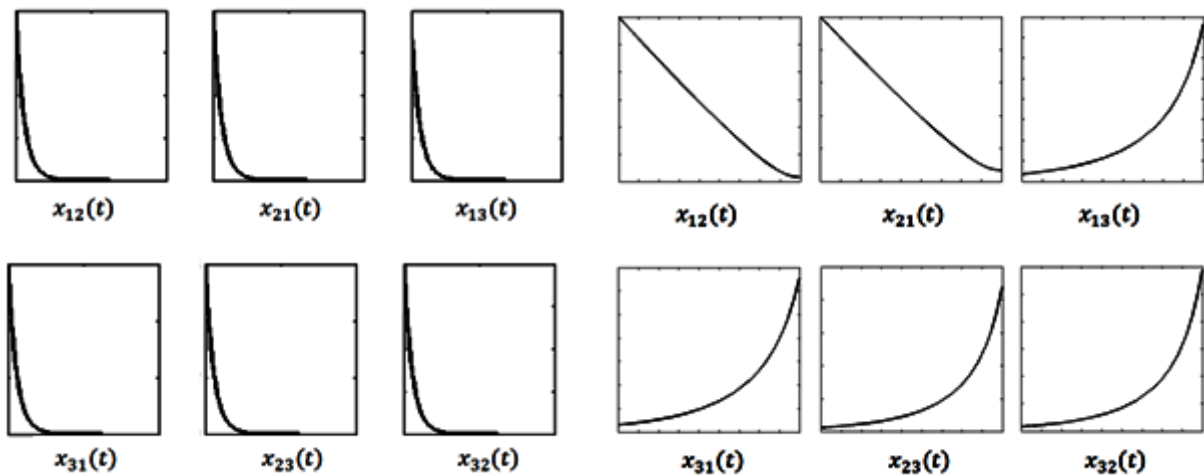
выполняются для всех $t: 0 < t \leq T$.

Смысл Теоремы 2 можно объяснить следующим образом: при выполнении условий действия теоремы, позиции и отношения участников друг к другу расходятся. Расходимость трактуется как невозможность в заданной системе участникам иметь близкие позиции при условии, что число переменных фиксировано (участники могут сблизиться только если в систему добавится дополнительный элемент, например, новый участник).

В разделе 1.3 проводится численное исследование устойчивых состояний модели в частных случаях

Было произведено численное моделирование с помощью метода Рунге-Кутты 4-го порядка решений системы (1)-(3) для различных начальных условий.

1. Показано, что если число участников (K) = 3, а число обсуждаемых вопросов (N) = 1, то возможны следующие стабильные состояния: все дистанции между участниками близки между собой; либо между двумя близки, а по отношению к третьему нет (см. Рис. 1).



Все дистанции близки

Между двумя дистанции близки, а по отношению к третьему нет

Рис. 1. Стабильные состояния при $K = 3, N = 1$

2. Если $K = 4$, $N = 1$, возможны следующие стабильные состояния: все дистанции между участниками близки, дистанции между тремя близки, дистанции в двух парах близки, дистанция близка только в одной паре.

Таким образом, в результате исследования модели показано соответствие традиционным моделям групповой динамики (с течением времени в группе устанавливаются отношения, не создающие диссонанса) и продемонстрирована возможность более двух устойчивых состояний (как и в подходе Дж. Дэвиса).

Во второй главе предлагается эмпирический метод измерения близости политических позиций, обеспечивающий насыщение модели исходными данными.

Заметим, что в системе (1)-(3), описывающей разрабатываемую модель, рассматриваются величины двух типов: 1) $x_{ij}(t)$ – субъективное расстояние от i -го участника до j -го; 2) разность $z_i^n(t) - z_j^n(t)$ – близость позиций между i -го и j -го участников по n -му вопросу.

Первая величина $x_{ij}(t)$ измеряется социометрическими и психологическими методами в достаточно хорошо проработанной научной области, которая развивается ещё с XIX века.

Вторая величина $z_i^n(t) - z_j^n(t)$ традиционно измеряется методами на основе контент-анализа, требующими привлечения экспертной оценки (например, в проекте “Manifesto”). В работе предлагается альтернативная методика определения близости политических позиций, заявленных в текстах.

Суть методики состоит в том, что при определении близости позиций анализируется синтагматическая близость текстов, выражающих эти позиции. Синтагматический подход к измерению расстояний в тексте – это понимание значения слова через контекст, в котором оно употреблено. При этом измерение расстояния происходит через сопоставление синтагматических свойств. Этот подход был реализован в методе латентно-семантического анализа (ЛСА) – запатентованный в 1988 году метод, предложенный американскими учёными¹⁷. ЛСА основан на «гипотезе о том, что между отдельными словами и обобщённым контекстом (предложениями, абзацами и целыми текстами), в которых они встречаются, существуют неявные (латентные) взаимосвязи, обуславливающие совокупность взаимных ограничений»¹⁸. Основное предположение состоит в том, что данные взаимосвязи являются различными в текстах, выражающих различные

¹⁷ T. Landauer, P.W. Foltz, D. Laham. Introduction to Latent Semantic Analysis. Discourse Processes 25: 259–284 (1998).

¹⁸ Митрофанова О.А. Семантические расстояния: проблемы и перспективы // XXXIV Международная филологическая конференция: Вып. 21. Прикладная и математическая лингвистика. СПб., 2005.

политические позиции. Так, нетрудно представить себе, что фамилия политического деятеля употребляется, как правило, в положительном контексте его сторонниками, и в негативном – противниками. Это относится не только к фамилиям и названиям партий, но также к отдельным политическим событиям, проектам и т.д. Тем самым, выражающие политические позиции тексты могут быть классифицированы путём выделения контекста, в который эти тексты погружают отдельные слова.

Основой для применения метода латентно-семантического анализа является выдвинутая автором гипотеза:

Гипотеза. *Близость политических позиций связана с синтагматической близостью текстов (фрагментов, образующих тексты), выражающих эти позиции.*

В отличие от традиционного способа применения ЛСА, проводится дополнительная препроцессинговая обработка текстов – лемматизация¹⁹ и удаление специально подобранных стоп-слов (слов, не несущих семантической нагрузки).

В Главе 3 приведено описание разработанного программного комплекса, а также представлены результаты его апробации на экспериментах по измерению близости позиций, представленных в предвыборных программах политических партий 2007 и 2011 гг.

В разделе 3.1 предложены результаты сопоставления различных политических текстов.

Разработан программный комплекс, использующий предложенный эмпирический метод численного измерения без привлечения эксперта близости политических позиций. С помощью разработанного программного комплекса были проанализированы предвыборные программы политических партий на выборах 2007 и 2011 годов в Государственную Думу.

Одним из способов представления результатов анализа является построение диаграмм, подобных изображенным на Рис. 2. Каждому фрагменту соответствуют одна строка и один столбец диаграммы. Ячейка, находящаяся на пересечении, например, 10-ой строки и 25-го столбца (так же, как ячейка 25-ой строки и 10-ого столбца) описывает степень синтагматической близости 10-го и 25-го фрагментов. При этом, чем более синтагматически близкими являются два фрагмента, тем более темным цветом закрашена ячейка.

На Рис. 2 фрагменты 1-15 принадлежат программе Единой России (ЕР), фрагменты 16-46 – программе партии «Яблоко». Темный квадрат, составленный из строк и столбцов 1-15, свидетельствует о высокой

¹⁹ Сегалович И. В. Как работают поисковые системы // Мир Интернет. – 2002. – №10.

синтагматической близости этих фрагментов, и в соответствии с предложенной гипотезой – о близости высказанных в этих фрагментах политических позиций. Диаграмма также показывает, что в программе Яблока выделяются две части, каждая из которых обладает более высокой внутренней связностью, чем программа в целом. Граница между ними приходится на фрагмент 30. Более подробное рассмотрение показало, что она соответствует границе разделов II «Альтернатива – социальное государство» и III «Благосостояние для всех» программы Яблока.

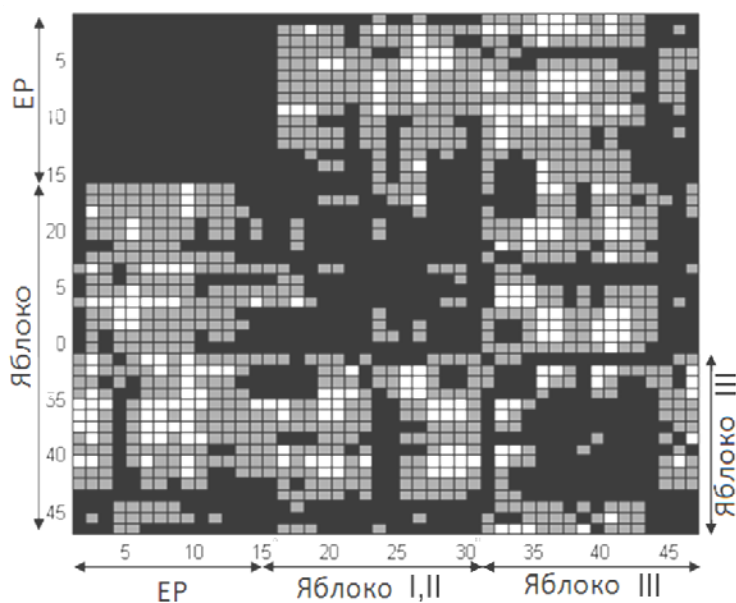


Рис. 2. Диаграмма сходства для программ Единой России и Яблока на выборах 2007 года

Программа ЕР (фрагменты 1-15) четко отделена от программы Яблока, в которой можно выделить две внутренне связные части, соответствующие разделам I,II (фрагменты 16-30) и разделу III (фрагменты 31-46).

Предвыборной программой в 2011 году на съезде Единой России было принято решение считать тексты выступлений Д.Медведева и В.Путина. Априори можно было бы предположить, что выступление Д.Медведева, имеющего репутацию либерала, окажется более близким программе Яблока, чем выступление В.Путина. Однако латентно-семантический анализ показывает (Рис. 3, показаны только строки, соответствующие программе Яблока), что это не так: пересечение столбцов 1-11, относящихся к тексту Медведева и строк 41-67, относящихся к партии «Яблоко» имеет более светлый тон, чем пересечение столбцов, относящихся к тексту Путина и строк, относящихся к тексту партии «Яблоко».

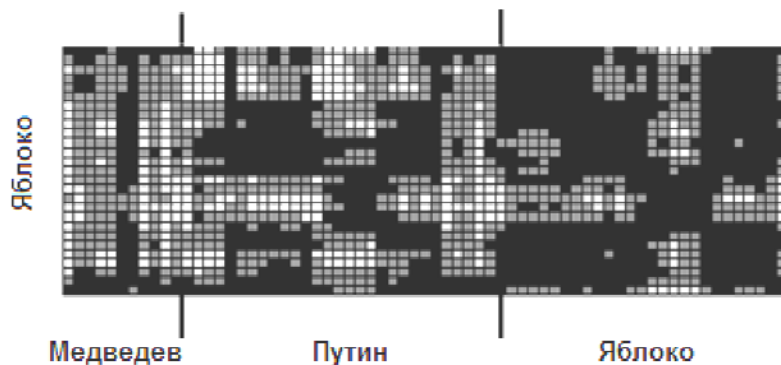


Рис. 3. Диаграмма сходства предвыборных программ ЕР и Яблока 2011 года (показаны только строки, соответствующие программе Яблока)

Сопоставление программ КПРФ и ЕР 2011 года показывает, в частности, что они слабо различимы, если рассматривать программу ЕР как целое. Если же (Рис. 4) рассмотреть корреляцию строк 1-26 (КПРФ) отдельно со столбцами 27-37 (Медведев) и 38-64 (Путин), то можно увидеть несходство программы КПРФ с выступлением Медведева, высокую близость с первой частью и несходство с последней частью выступления Путина.



Рис. 4. Диаграмма сходства предвыборных программ КПРФ и ЕР 2011 года (показаны только строки, соответствующие программе КПРФ)

В разделе 3.2 вводится понятие числовой меры синтагматической близости программ и проводится кластеризация текстов предвыборных программ партий в Государственную Думу в 2007 и 2011 годах.

Обозначим синтагматическое расстояние между i -м и j -м фрагментами через r_{ij} , введём величины

$$A = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_1} \left(\frac{1+r_{ij}}{2} \right) + \sum_{i=N_1+1}^{N_1+N_2} \sum_{j=N_1+1}^{N_1+N_2} \left(\frac{1+r_{ij}}{2} \right)}{N_1^2 + N_2^2} \quad (12),$$

$$B = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=N_1+1}^{N_1+N_2} \left(\frac{1+r_{ij}}{2} \right) \quad (13).$$

Определение 2. Мерой синтагматической близости программ будем называть отношение $R = 200B / A$.

Вычисленные таким образом значения R для предвыборных партийных

программ четырех российских партий на выборах 2007 и 2011 годов сведены в Табл. 1 и Табл. 2.

Табл. 1. Синтагматическая близость партийных программ 2007 года

| 2007 | ЕР | КПРФ | ЛДПР | Яблоко |
|--------|----|------|------|--------|
| ЕР | 0 | 99 | 92 | 87 |
| КПРФ | 99 | 0 | 95 | 91 |
| ЛДПР | 92 | 95 | 0 | 91 |
| Яблоко | 87 | 91 | 91 | 0 |

Табл. 2. Синтагматическая близость партийных программ 2011 года

| 2011 | ЕР | КПРФ | ЛДПР | Яблоко |
|--------|----|------|------|--------|
| ЕР | 0 | 87 | 91 | 89 |
| КПРФ | 87 | 0 | 96 | 85 |
| ЛДПР | 91 | 96 | 0 | 90 |
| Яблоко | 89 | 85 | 90 | 0 |

С помощью метода иерархической кластеризации проанализированы данные Таблицы 1 в отношении того, какие партийные программы наиболее близки и далеки друг от друга. Результаты представлены на Рис. 5.

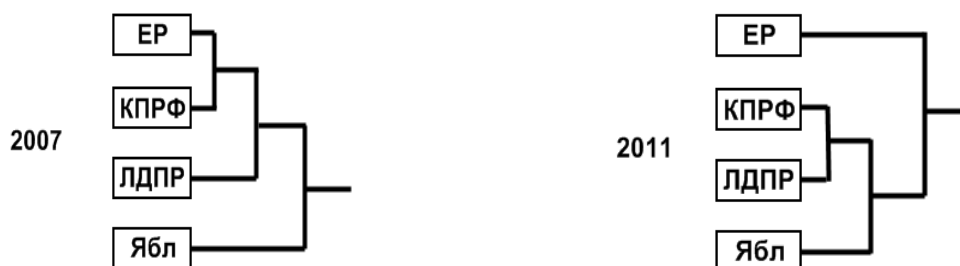


Рис. 5. Структура синтагматической близости программ 2007 и 2011 годов

Таким образом, в предвыборную кампанию 2007 года наиболее близкими были программы ЕР и КПРФ, а наиболее отдаленной от других – программа Яблока. В кампанию 2011 года произошла определенная перестройка: наиболее близкими были программы КПРФ и ЛДПР, а наиболее отдаленной от других – программа ЕР. В целом, в 2011 г. программы различных партий стали более похожими друг на друга, чем в 2007.

В заключении подводятся итоги диссертационной работы и формулируются основные результаты.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах, рекомендованных ВАК

1. *Корнилина Е.Д.* Применение латентно-семантического анализа к определению позиции политических блогеров // М.: Социология, Издательский дом «Граница». – 2011. – №1. – С. 162-164.
2. *Корнилина Е.Д., Петров А.П.* Динамическая модель близости позиций пользователей социальных сетей // М.: Математическое моделирование, Академиздатцентр «Наука». – 2012. – Т. 24. – №10. – С. 89-97.
3. *Корнилина Е.Д., Петров А.П.* Латентно-семантический анализ предвыборных партийных программ на выборах в Государственную Думу 2007 и 2011 годов // М.: Вестник МГУ. Сер. 12: Политические науки, Издательство МГУ. – 2013. – №2. – С. 90-98.

Прочие публикации

4. *E. Kornilina, A. Petrov* The investigation of political blogs by means of latent semantic analysis // Труды Третьей международной конференции «Математическое моделирование социальной и экономической динамики» (MMSED-2010). 23-25 июня 2010г. / Науч. ред. М.Г. Дмитриев, А.П. Петров, Н.П. Третьяков. – М.: Ленанд. – 2010. – Р. 142-143.
5. *E. Kornilina, A. Petrov* Research of political blogs by means of LSA // Интеллектуализация обработки информации: 8-я международная конференция. Республика Кипр, г. Пафос, 17-24 октября 2010г.: Сборник докладов. – М.: МАКС Пресс. – 2010. – С. 508-511.
6. *Корнилина Е.Д.* Выявление степени коррелируемости позиций политических блогеров // Стратегия инновационного развития России как особой цивилизации в XXI веке: VI Всероссийская научная конференция «Сорокинские чтения – 2010»: Сборник тезисов. – М.: МАКС Пресс. – 2010. – С. 758-760.
7. *Корнилина Е.Д., Петров А.П.* О приложении латентно-семантического анализа к исследованию текстов социально-политической тематики // Теория активных систем. Труды международной научно-практической конференции (14-16 ноября 2011 г., Москва, Россия). Общая редакция – В.Н. Бурков, Д.А. Новиков. // М.: ИПУ РАН – 2011. – Т.2. – С. 266-269.
8. *Корнилина Е.Д., Петров А.П.* Модель динамики близости политических позиций // Теория активных систем. Труды международной научно-практической конференции (14-16 ноября 2011 г., Москва, Россия). Общая редакция – В.Н. Бурков, Д.А. Новиков. // М.: ИПУ РАН – 2011. – Т.2. – С. 270-272.

9. *Корнилина Е.Д.* Моделирование изменения отношений в политических группах // Современные проблемы математического моделирования. Сборник трудов XIV молодежной конференции-школы с международным участием // Ростов-на-Дону, издательство Южного федерального университета. – 2011. – С.188-192.
10. *Корнилина Е.Д., Петров А.П.* О динамике близости политических позиций: модель и методика измерения // Математическое моделирование социальных процессов. Выпуск 12-13: Сборник статей / Под ред. А.П. Михайлова. — М.: Издательство «Спутник +». – 2012. – С.42-51.
11. *Петров А.П., Корнилина Е.Д.* Исследование близости политических позиций методом латентно-семантического анализа // Сборник докладов XII Международной научной конференции по проблемам развития экономики и общества. Москва. – Издательский дом ВШЭ. – 2012. – Т. 2. – С.334-342.
12. *Корнилина Е.Д., Петров А.П.* Аналитическое и численное исследование модели динамики позиций членов замкнутой группы // М.: Кавказские научные записки, Издательство Российского государственного торгово-экономического университета. – 2012. – №2. – С. 227-235.
13. *Петров А.П., Корнилина Е.Д.* Сравнительное исследование предвыборных партийных программ 2007 и 2011 гг. на основе латентно-семантического анализа // Сборник докладов XIII Международной научной конференции по проблемам развития экономики и общества. Москва. – Издательский дом ВШЭ. – 2012. – Т. 4. – С. 96-105.
14. *Корнилина Е.Д.* Анализ изменения высказываемых позиций на примере модели взаимодействующих в замкнутой группе акторов. // Интеллектуализация обработки информации: 9-я международная конференция. Черногория, г. Будва, 2012 г.: Сборник докладов. – М.: Торус Пресс. – 2012. – С. 662-664.