

УДК 007.5, 004.8

ПОИСК СОГЛАСОВАННЫХ РЕШЕНИЙ ЧЕЛОВЕКОМ И РОБОТОМ. ПРОБЛЕМА МАНИПУЛИРОВАНИЯ

В.Н. Жидков (*vladimir_zhidkov@mail.ru*)
Московский авиационный институт, Москва

Н.В. Ким (*nkim2011@list.ru*)
Московский авиационный институт, Москва

Аннотация. В статье рассматриваются вопросы общения человека и робота с целью выработки согласованных решений. Определена структура обработки информации робота, обеспечивающая эффективное принятие согласованных решений. Показано, что при решении различных целевых задач человеком и роботом, робот может реализовать манипулирование человеком.

Ключевые слова: человек, психоэмоциональное состояние, психологическое воздействие, робот, целевые задачи, принятие решение, правила манипулирования.

Введение

В рамках данной статьи будем считать, что манипулированием является достижение своих собственных целей *манипулятором* за счет психологического воздействия на конкретного человека.

В будущем дальнейшее развитие робототехники приведет к широкому распространению различных роботов (Р): водителей, продавцов, строителей и пр., к которым человек (Ч) будет обращаться за помощью. При этом в процессе общений Ч и Р возможно возникновение коллизий, связанных с недостаточным пониманием Ч и Р, различием решаемых целевых задач и т.д. Если основные психоэмоциональные особенности людей в ближайшие десятилетия вряд ли будут существенно меняться, то роботы могут регулярно оснащаться все более совершенными программно-аппаратными средствами, обеспечивающими более эффективное выполнение требуемых правил поведения.

С учетом данного фактора актуальным представляется вопрос о том, как должен быть оснащен Р, чтобы наиболее эффективно находить совместно с Ч согласованные решения. Если человек при выполнении каких-либо задач

будет обращаться за помощью к роботам, имеющим другие целевые задачи, то и человек, и робот должны будут принимать некоторые, устраивающие обоих, согласованные решения. При этом предполагается, что каждый из них будет стремиться в наибольшей степени обеспечить выполнение собственной целевой задачи.

Термин «согласованность» подразумевает, что Ч в процессе общения может пойти на некоторые уступки, т.е. изменить свои исходные цели с учетом имеющихся целевых функций Р. Таким образом, Р, целенаправленно изменяя исходные цели Ч с учетом своих интересов, является *манипулятором*.

1. Концепция построения комплексного алгоритма

Рассмотрим структуру комплексного алгоритма «**Принятие согласованных решений**», позволяющего Р «договариваться» с Ч о совместных действиях, а также манипулировать человеком.

В «базовом» случае предполагается, что цели Ч и Р совпадают. Тогда в Р может быть реализован вариант, в котором после получения запроса (Q) от Ч формируется набор возможных совместных действий $U(Q)$. Если Ч выбирает некоторый $u_i \in U(Q)$, то считается, что согласованное решение принято.

Если цели Ч и Р не совпадают, то необходимо находить компромиссное решение, например, с использованием аппарата теории игр.

Одним из возможных вариантов нахождения согласованного решения является использование Р методов манипулирования. Наиболее просто реализуется вариант аналогичный классическому «программному» управлению. Здесь робот предварительно оснащается *набором приемов психологического воздействия (манипулирования) (ППВ)*, которые могут им использоваться. Существует большое количество подобных приемов [1] и их программно-алгоритмическая реализация вполне возможна.

Рассмотрим ситуацию, когда робот-водитель такси с целью повышения прибыли пытается уговорить пассажира выбрать более длинный маршрут (к заказанному пункту назначения), чем маршрут $M(H)$, выбранный «Навигатором». В этом случае может, например, использоваться прием «*Навязывание выбора*».

Алгоритм реализации приема состоит в следующем:

Из имеющегося набора возможных маршрутов выбираются маршруты $M(1)$ и $M(2)$ более длинные, чем $M(H)$, и проходящие через какие-либо достопримечательности. Пассажиру предлагается: «Если хотите увидеть уникальное здание и площадь, то поедem маршрутом $M(1)$, а если: парк и озеро, то $M(2)$ ». На подобное предложение некоторые из пассажиров

выберут один из предлагаемых вариантов, что позволяет достичь Р своей цели.

Эффективность воздействия приемов манипулирования на различных людей различна. В существенной степени она зависит от психотипа конкретного человека и его эмоционального состояния [Бодалев, 1995].

Поэтому в следующем варианте более эффективного манипулирования учитывается психотип человека.

Известны методы определения психотипов по внешним признакам. Например, в [3] выделяются признаки: "сутулый, выпяченный живот, худой, интуиция +1, логика +1, экстраверсия +1" или "неулыбчив, сжатые губы, сухое выражение лица - логика +1", по которым делается заключение о психотипе данного Ч. Подобные признаки могут быть выделены с использованием известных технологий компьютерного зрения на базе различных методов распознавания [Kim N и др., 2014].

Данные о психотипе человека должны собираться и храниться в психологическом паспорте человека (ППЧ(n)), где $n = 1, \dots, N$ индекс конкретного человека. Эти данные позволят формировать соответствующие индивидуальные ППВ(n).

Известно, что решения, который принимает Ч в существенной степени зависят от его эмоционального состояния [[Бодалев, 1995], поэтому логично в процессе манипулирования учитывать и этот фактор.

Для оценки состояния человека робот должен быть *оснащен модулем оценки эмоционального состояния Ч*.

В настоящее время уже разработаны программы, которые позволяют автоматически оценивать эмоциональное состояния людей на основе анализа изображений их лиц. Например, система видео-аналитики «Монитор Эмоций» [5] (бесплатная версия «ЭмоСкоп») позволяет в режиме реального времени контролировать **эмоциональный фон** и другие характеристики человека по выражению лица с использованием экспертной системы.

Учет эмоционального состояния позволит использовать более эффективные ППВ(n, e, t), где e – индекс эмоционального состояния, t – текущий момент времени.

Таким образом, в ППЧ необходимо вносить сведения о психотипе и эмоциональном состоянии Ч (в соответствии с его реакцией на определенные раздражители).

2. Модель взаимодействия Ч и Р

Рассматриваемая модель взаимодействия робота и человека разработана в среде MATLAB с использованием подсистемы моделирования динамических процессов SIMULINK. Модель разработана

в соответствии с принципами построения адаптивных психофизических интерфейсов, способных подстраиваться под конкретного человека-оператора и учитывающих в процессе общения его эмоциональное состояние [Жидков и др., 1999], [Жидков, 2013].

Процесс общения человека и робота рассматривается как непрерывный процесс устранения рассогласования между текущим условным состоянием некоторой потребности человека, и желаемым (мотивированным) состоянием этой потребности.

В модели условное состояние потребности человека кодируется некоторым числовым значением, абсолютная величина которого не несет никакой смысловой нагрузки - важна степень рассогласования между текущим и целевым значениями этой потребности (рассогласование целевой функции). Целевое значение потребности (целевая функция) определяется мотивацией человека, но робот может пытаться изменить ее по своему усмотрению (см. выше).

Модель реализована на основе нейро-нечетких систем ANFIS [Жидков, 2013], удобный инструмент для проектирования которых предлагается в пакете расширения MATLAB - Fuzzy Logic Toolbox.

Нейро-нечеткая система имеет базу знаний, содержащую нечеткие правила, формируемые экспертом, в которых заключен основной интеллект системы.

Кроме того, нейро-нечеткая система ANFIS способна обучаться на примерах решения задачи, которые могут быть получены от эксперта либо самостоятельно роботом в процессе адаптации его к окружающему миру.

Разработанная модель взаимодействия робота и человека содержит три контура:

1. Контур динамики общения. Содержит модуль человека (рис.1) и модуль "прикладного" интеллекта, который необходим роботу для решения целевой задачи (рис.2). Модуль человека содержит две нейро-нечеткие системы - ANFIS 1 и ANFIS 2. В базах знаний этих систем в виде нечетких продукций заложена информация о психотипе конкретного человека, а также информация, необходимая для моделирования эмоционального состояния человека в процессе общения с роботом.

Система ANFIS 1 имеет четыре входа, на основе которых определяется выходной параметр системы - целевая функция человека:

1. Целевой стимул (мотивация);
2. Мешающие факторы (шум);
3. Сигнал с выхода Out 1 модуля управления поведением человека (см. ниже);
4. Время общения с роботом.

Система ANFIS 2 предназначена для моделирования психотипа человека и фактически регулирует коэффициент демпфирования в контуре динамики общения. Система также имеет четыре входа:

1. Начальное демпфирование;
2. Время общения с роботом;
3. Информация от робота (In1), на который поступает сигнал с выхода Out 1 модуля управления эмоциями человека (см. ниже);
4. Текущее рассогласование целевой функции.

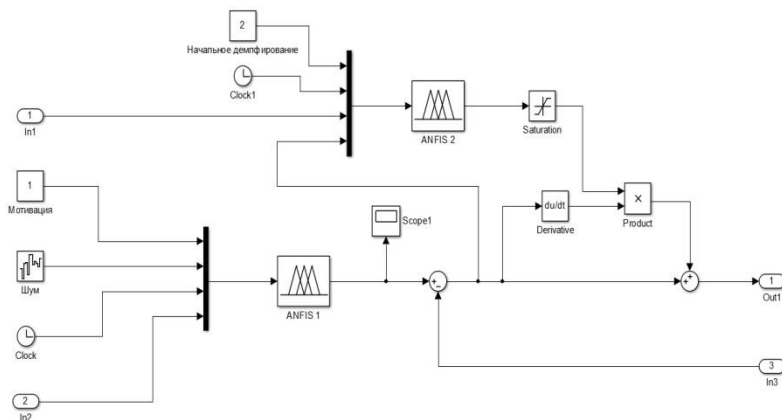


Рис. 1. Модель человека

Модуль прикладного интеллекта робота (рис.2) содержит нейро-нечеткую систему ANFIS3, в которой в виде нечетких продукций заложены знания, необходимые для устранения рассогласования между текущим и целевым состоянием человека и/или робота. Сигнал In1 модуля поступает сигнал с выхода Out1 модели человека (демпфированное рассогласование целевой функции). Этот сигнал дважды интегрируется (идет накопление информации) и затем поступает на "лингвистический" регулятор, реализованный на нейро-нечеткой системе, которая формирует информационный сигнал человеку (выход Out1), необходимый для удовлетворения текущей условной потребности человека, - сигнал обратной связи. С выхода Out1 сигнал обратной связи поступает на вход In3 модели человека - вход прямого восприятия информации от робота.

На вход блока ANFIS 3 поступают 3 сигнала:

1. Однократно проинтегрированный сигнал о рассогласовании целевой функции;

2. Двухкратно проинтегрированный сигнал о рассогласовании целевой функции;
3. Мотивация робота, которая задается внешней системой управления роботом и может отличаться от мотивации человека. Сигнал мотивации робота с выхода Out2 также подается в модуль управления мотивацией человека (см. ниже).

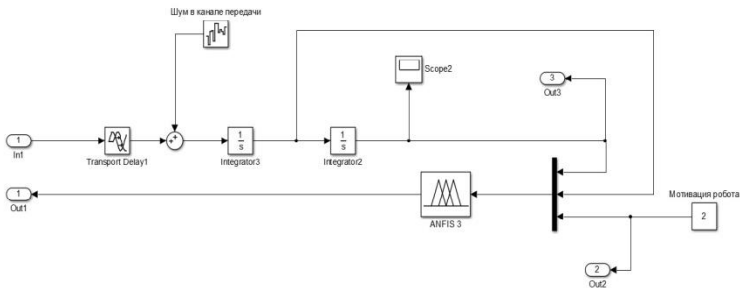


Рис. 2 Модель прикладного интеллекта

2.Контур управления эмоциональным состоянием человека. Содержит модуль определения диссонанса (рис.3) и модуль управления эмоциями человека (рис.4). Здесь принято, что *диссонанс* является обобщенной характеристикой текущего психоэмоционального состояния человека и определяется как функция текущего рассогласования целевой функции, тенденции ее изменения во времени, а также времени общения человека и робота.

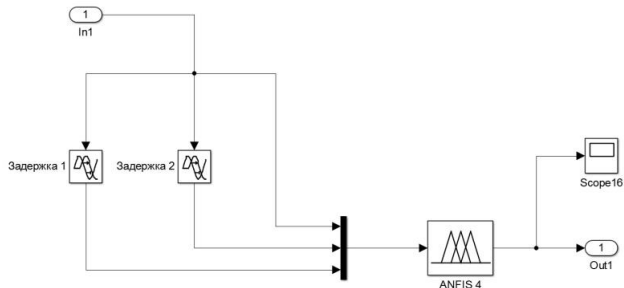


Рис. 3. Модель определения диссонанса

Нечеткие лингвистические правила определения диссонанса заложены в базу знаний нейро-нечеткой системы ANFIS 4.

На вход системы In1 с выхода Out 3 модуля прикладного интеллекта поступает двукратно проинтегрированный сигнал текущего рассогласования целевой функции;

Выход модуля Out1 непосредственно связан со входом In3 модуля управления эмоциями человека.

Модуль управления эмоциями человека (рис.4) реализован на нейро-нечеткой системе ANFIS5, которая содержит как базу знаний с нечеткими лингвистическими правилами коррекции психотипа человека, сформированными на основе общих знаний о методах управления эмоциями человека (ППВ), так и базу данных, содержащую индивидуальные психоэмоциональные паспорта конкретных людей (ППЧ(n)). Выход системы Out1 поступает на вход In1 системы ANFIS 2, непосредственно моделирующей психотип человека.

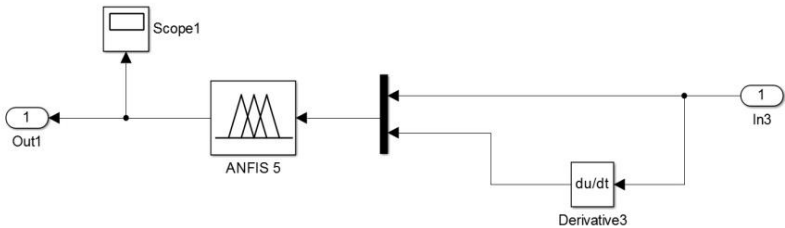


Рис. 4. Модель управления эмоциями человека

3.Контур воздействия на мотивацию человека. Модуль управления целевой функцией человека (рис.5) реализован на нейро-нечеткой системе ANFIS6, которая содержит как базу знаний с нечеткими лингвистическими правилами психологического воздействия (манипулирования) (ППВ) на человека с целью принудить его к изменению целевой функции (см. выше), так и базу данных, содержащую индивидуальные психоэмоциональные паспорта конкретных людей.

База знаний и база данных наполняется экспертными знаниями, и пополняется в процессе функционирования и самообучения робота. На вход системы In1 поступает информация о текущем диссонансе с выхода Out1 модуля определения диссонанса (рис.3) а также сигнал мотивации робота (вход In3) с выхода Out2 модуля прикладного интеллекта (рис.2).

Выход системы Out1 подается на вход восприятия информации от робота (In2) модуля человека (рис.1).

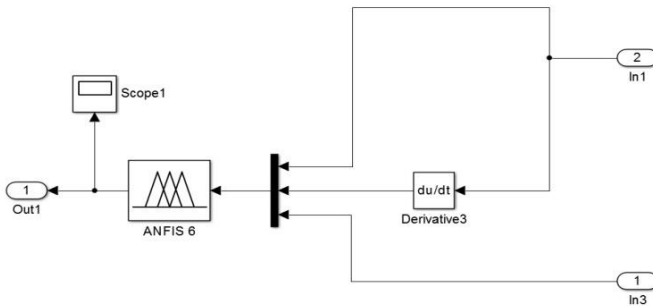


Рис. 5. Модель управления целевой функцией человека

3. Результаты моделирования

На рис. 6 и рис. 7 представлены результаты моделирования процесса общения человека и робота. В данном случае (ситуация 1) робот только помогает устранить рассогласование условной потребности человека (P). Исходное состояние $P=0$, желаемое состояние $P=1$. На рис.6 показан процесс согласования потребности, на рис. 7 - диссонанс.

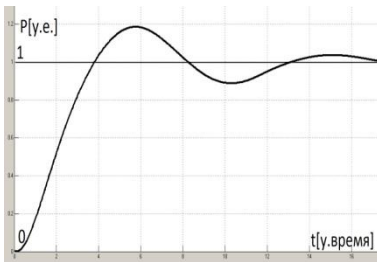


Рис. 6. Процесс согласования 1

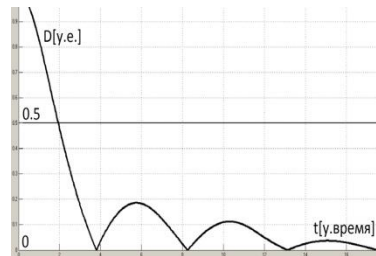


Рис. 7. Диссонанс согласования 1

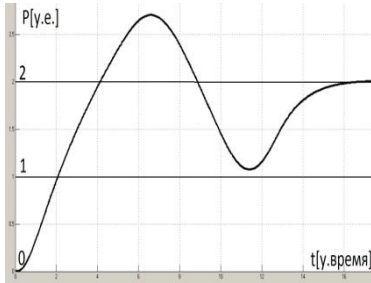


Рис. 8. Процесс согласования 2



Рис. 9. Диссонанс согласования 2

Было проведено моделирование различных ситуаций, в частности, следующий сценарий общения человека и робота (рис. 6):

1. Робот предлагает вариант действий u_1 ,
2. На $t \geq 4$ человек принимает решение, что действие слишком рискованное,
3. На $t \geq 6$ робот предлагает вариант u_2 ,
4. На $t \geq 8$ человек принимает решение, что действие недостаточно активно,
5. На $t \geq 10$ робот предлагает вариант u_3 ,
6. На $t \geq 14$ человек принимает решение, что вариант удовлетворителен (рассогласование $< 0,5$).

На рис. 8 и рис. 9 представлены результаты общения в ситуации (ситуация 2), когда робот навязывает свою мотивацию. Исходное состояние человека $P=0$, желаемое состояние человека $P=1$, мотивация робота $P=2$. Видно, что процесс идет с большими колебаниями диссонанса, человек сопротивляется мотивации робота, но в конце концов робот убеждает человека принять его (робота) мотивацию.

Заключение

1. Показано, что возможно создание систем ИИ, которые смогут манипулировать человеком. При этом эффективность этих систем существенно возрастет, если на отдельных людей будут формироваться их индивидуальные психоэмоциональные паспорта (ППЧ);
2. Определена структура комплексного алгоритма «Принятие согласованных решений»;
3. Проведено моделирование процесса общения человека и робота, показавшее качественное сходство с реальным процессом;

4. Авторы считают, что использование подобных систем (особенно в сетевых технологиях) опасно с точки зрения возможности создания единых ППЧ и распространения технологий манипулирования людьми.

Список литературы

- [Бодалев, 1995] Бодалев А. А. Личность и общение. — М.: Международная педагогическая академия, 1995. — 328 с. <https://bbf.ru/magazine/2/2746/>
- [Kim N и др., 2014] Kim N., Bodunkov N. «Computer Vision in Advanced Control Systems: Innovations in Practice», Volume 2, Editors M. Favorskaya, Lakhmi C. Jain, Springer 2014. — 295 p. <http://psychotype.ru/article/a-15.html>
- [Жидков, 2013] Жидков В.Н. Мягкие вычисления в интегрированных системах летательных аппаратов: Учеб.пособие.- М.: Изд-во МАИ, 2013.- 388с. <http://911.prolan.ru/applying/emotionsmonitor.html>
- [Жидков и др., 1999] Жидков В.Н., Жидков М.В, Красильщиков М.Н. Обучаемый интеллектуальный помощник пилота. Анализ вариантов реализации // Полет. 1999. №4