Модели формирования многомерных мнений с ограниченным доверием

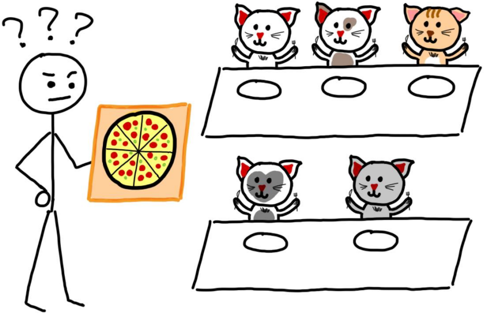
Забарянская Ирина

Московский физико-технический институт
кафедра интеллектуальных систем

Научно-исследовательская работа
21 декабря 2024 года

Многомерные мнения и ограниченное доверие

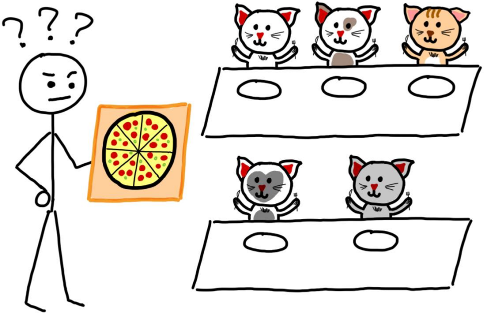
Многомерное мнение: набор мнений по различным темам.
Например: как распределить ресурс



Ограниченное доверие: склонность индивидуума доверять только мнениям, которые близки к его собственному.
Близость – ?

Многомерные мнения и ограниченное доверие

Многомерное мнение: набор мнений по различным темам.
Например: как распределить ресурс

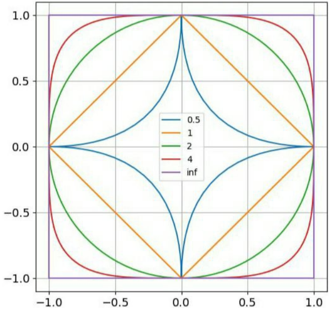


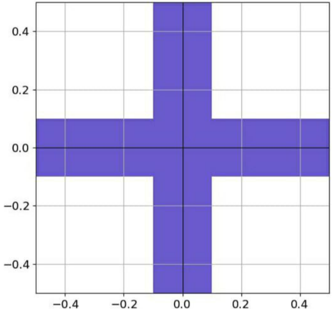
Ограниченное доверие: склонность индивидуума доверять только мнениям, которые близки к его собственному.
Близость – ?

Модель

1. Множество агентов $\mathcal{V} \triangleq \{1, \dots, n\}$
2. Дискретное время $t = 0, 1, \dots$
3. Мнение агента $i \in \mathcal{V}$ в момент времени t это вектор $\xi^i(t) = (\xi^i_1(t), \dots, \xi^i_d(t)) \in \mathbb{R}^d$
4. "Близость" мнений ξ и ξ' характеризуется доверительным множеством $\mathcal{O} \subseteq \mathbb{R}^d$: $\xi' - \xi \in \mathcal{O}$

Примеры доверительных множеств (слева направо): шары в ℓ_p метрике, неограниченное множество $\mathcal{O} = \{\xi : \min_{i\in\{1,2\}} |\xi_i| \le 0.1\}$.



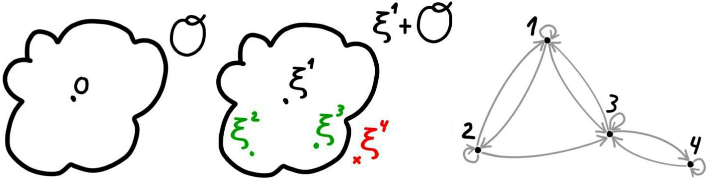


Граф доверия

Обозначение: $\boldsymbol{\xi} \triangleq (\xi_k^i)_{k=1,\dots,d}^{i\in\mathcal{V}}$ – семейство мнений.

Агент i принимает во внимание мнения агентов из множества $\mathcal{N}_i(\boldsymbol{\xi}) \triangleq \{j : \xi^j \in \xi^i + \mathcal{O}\}$.

Предположение: $0 \in \mathcal{O}$, то есть агент доверяет собственному мнению: $i \in \mathcal{N}_i(\boldsymbol{\xi})$.



Граф доверия $\mathcal{G}(\boldsymbol{\xi}) \triangleq (\mathcal{V}, \mathcal{E}(\boldsymbol{\xi}))$:

1. Вершины = агенты
2. Ребра: $i \to j$, если $j \in \mathcal{N}_i(\boldsymbol{\xi})$ (агент i доверяет агенту j)

Система A (generalized Hegselmann-Krause)

Агенты синхронно обновляют свои мнения.

Новое мнение агента i – среднее “близких” мнений:

$$\xi^i(t+1) = \frac{1}{|\mathcal{N}_i(\boldsymbol{\xi}(t))|} \sum_{j\in\mathcal{N}_i(\boldsymbol{\xi}(t))} \xi^j(t), \quad i \in \mathcal{V} \quad \text{(A)}$$

Динамическая система с разрывной правой частью!

Решение может сходиться к неравновесной точке!

Hegselmann R., Krause U. Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation // Journal of Artificial Societies and Social Simulation. 2002.

Система A (generalized Hegselmann-Krause)

Агенты синхронно обновляют свои мнения.

Новое мнение агента i – среднее "близких" мнений:

$$\xi^i(t+1) = \frac{1}{|\mathcal{N}_i(\boldsymbol{\xi}(t))|} \sum_{j \in \mathcal{N}_i(\boldsymbol{\xi}(t))} \xi^j(t), \quad i \in \mathcal{V} \quad \text{(A)}$$

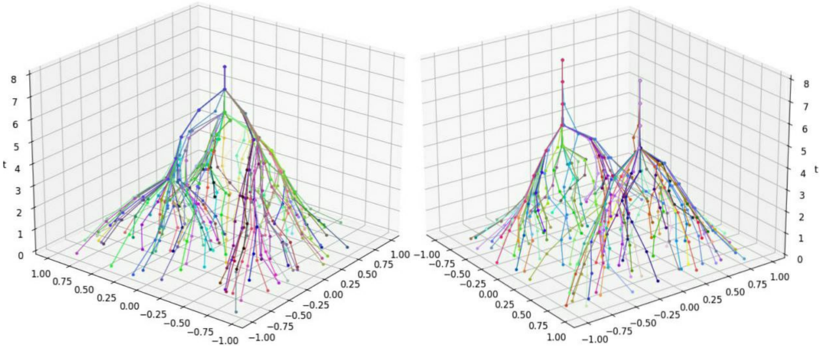
Динамическая система с разрывной правой частью!

Решение может сходиться к неравновесной точке!

Hegselmann R., Krause U. Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation // Journal of Artificial Societies and Social Simulation. 2002.

Консенсус и рассогласование

Начальные $2d$**-мнения**: сгенерированы случайно из $[-1,1]\times[-1,1]$.
Доверительное множество: $\mathcal{O}=\{\xi : \|\xi\|_2 \le 0.6\}$.



Сходимость за конечное число шагов!

Сходимость мнений

Дополнительные предположения:

A1. Доверительное множество симметрично: $\mathcal{O} = -\mathcal{O}$.

A2. $0 \in \mathbb{R}^d$ внутренняя точка $\mathcal{O}$.

Теорема (о сходимости мнений при всех $\xi(0)$)

Из A1 следует:

- $\exists \xi^i(\infty) \triangleq \lim_{t\to\infty} \xi^i(t) \in \mathbb{R}^d, \quad \forall i \in \mathcal{V}$

Из A1 $\wedge$ A2 следует:

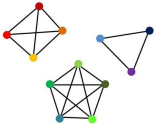
- $\xi(\infty)$ – равновесие (A)
- мнения сходятся за конечное число шагов

Классификация положений равновесия

Кластер агента i: множество $\{j \in \mathcal{V} : \xi^i = \xi^j\}$.

Семейство мнений $\boldsymbol{\xi}$ **кластеризовано**, если агенты из разных кластеров не доверяют друг другу: $\xi^i = \xi^j$ или $\xi^j - \xi^i \notin \mathcal{O}$.

Граф доверия разбивается на клики.



Теорема (классификация положений равновесия)

Если $\boldsymbol{\xi}^*$ кластеризовано, то $\boldsymbol{\xi}^*$ – равновесие (A).

Если $\mathcal{O}$ замкнуто, то любое кластеризованное равновесие устойчиво по Ляпунову.

Если $\mathcal{O} = -\mathcal{O}$, то все равновесия кластеризованы.

Система B (инерциальная)

В модели (A) агент i присваивает своему мнению вес $\frac{1}{|\mathcal{N}_i(\boldsymbol{\xi})|}$.

Вес собственного мнения может быть $\frac{1}{n}$ (зависит от графа).

Инерциальная модель (вес своего мнения не меньше константы):

$$\xi^i(t+1) = (1-h_i)\xi^i(t) + \frac{h_i}{|\mathcal{N}_i(\boldsymbol{\xi}(t))|} \sum_{j \in \mathcal{N}_i(\boldsymbol{\xi}(t))} \xi^j(t), \quad i \in \mathcal{V} \quad \text{(B)}$$

1. "Упрямый" агент: $h_i = 0$ (не рассматриваем)
2. Модель (A): $h_i = 1, \quad \forall i \in \mathcal{V}$
3. Без "упрямых" агентов равновесия (A) и (B) совпадают

B. Chazelle and C. Wang, Inertial Hegselmann–Krause systems // IEEE Trans. Autom. Control. 2017.

Система B (инерциальная)

В модели (A) агент i присваивает своему мнению вес $\frac{1}{|\mathcal{N}_i(\boldsymbol{\xi})|}$.

Вес собственного мнения может быть $\frac{1}{n}$ (зависит от графа).

Инерциальная модель (вес своего мнения не меньше константы):

$$\xi^i(t+1) = (1-h_i)\xi^i(t) + \frac{h_i}{|\mathcal{N}_i(\boldsymbol{\xi}(t))|} \sum_{j \in \mathcal{N}_i(\boldsymbol{\xi}(t))} \xi^j(t), \quad i \in \mathcal{V} \quad \text{(B)}$$

1. "Упрямый" агент: $h_i = 0$ (не рассматриваем)
2. Модель (A): $h_i = 1, \quad \forall i \in \mathcal{V}$
3. Без "упрямых" агентов равновесия (A) и (B) совпадают

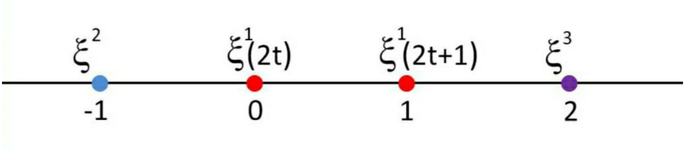
B. Chazelle and C. Wang, Inertial Hegselmann–Krause systems // IEEE Trans. Autom. Control. 2017.

Упрямые агенты: периодическая траектория

С упрямыми агентами мнения могут расходиться!

Пример

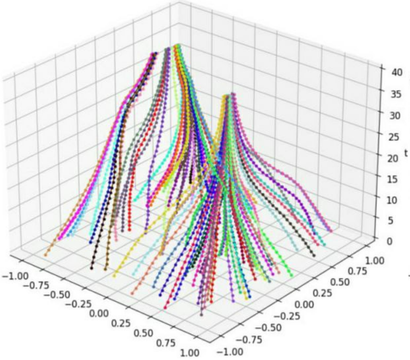
1. $n = 3$ агентов
2. Начальные мнения: $\xi^1(0) = 0$, $\xi^2(0) = -1$, $\xi^3(0) = 2$
3. Доверительное множество: $\mathcal{O} = (-3, 3) \setminus \{-1, 1\}$
4. Коэффициенты инерции: $h_1 = 1$, $h_2 = h_3 = 0$



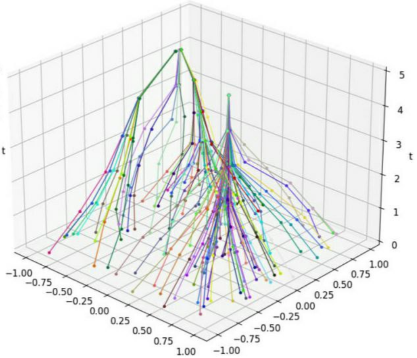
Мнение агента 1: $\xi^1(t) = (t \mod 2)$.

Инерция и скорость сходимости

Начальные $2d$**-мнения**: сгенерированы случайно из $[-1, 1] \times [-1, 1]$.
Доверительное множество: $\mathcal{O} = \{\xi : \|\xi\|_2 \leq 0.6\}$.



$h_i = h = 0.1$



$h_i = h = 0.9$

Асимптотическая сходимость. Скорость зависит от h!

Сходимость мнений

Дополнительные предположения:

A1. Доверительное множество симметрично: $\mathcal{O} = -\mathcal{O}$.

A2. $0 \in \mathbb{R}^d$ внутренняя точка $\mathcal{O}$.

Теорема (о сходимости мнений для всех $\boldsymbol{\xi}(0)$)

Из A1 следует:

- $\exists \xi^i(\infty) \triangleq \lim_{t\to\infty} \xi^i(t) \in \mathbb{R}^d, \quad \forall i \in \mathcal{V}$

Из A1 $\wedge$ A2 следует:

- граф перестает меняться через конечное время
- экспоненциально быстрая сходимость мнений

Если $\mathcal{O}$ открыто и A1 $\wedge$ A2, то

- $\boldsymbol{\xi}(\infty)$ – равновесие (B)

Система C (с упрямыми агентами)

Модель (C) = Модель (A) + упрямые агенты

1. Упрямый агент не меняет свое мнение:

$$\mathcal{N}_i(\boldsymbol{\xi}) \triangleq \begin{cases} \{j : \xi^j \in \xi^i + \mathcal{O}\}, & \text{если } i \text{ обычный агент} \\ \{i\}, & \text{если } i \text{ упрямый агент} \end{cases}$$

2. Динамическая система:

$$\xi^i(t+1) = \frac{1}{|\mathcal{N}_i(\boldsymbol{\xi}(t))|} \sum_{j \in \mathcal{N}_i(\boldsymbol{\xi}(t))} \xi^j(t), \quad i \in \mathcal{V} \quad \text{(C)}$$

Q1: Iryna Zabarianska and Anton V. Proskurnikov. Opinion Dynamics with Set-Based Confidence: Convergence Criteria and Periodic Solutions // IEEE Control Systems Letters. 2024.

Сходимость мнений

Дополнительные предположения:

A1. Доверительное множество симметрично: $\mathcal{O} = -\mathcal{O}$.

A2. $0 \in \mathbb{R}^d$ внутренняя точка $\mathcal{O}$.

A3. Упрямые агенты имеют одинаковое мнение ξ^*.

Теорема (о сходимости мнений при всех $\boldsymbol{\xi}(0)$)

Из A1 $\wedge$ A3 следует:

- $\exists \xi^i(\infty) \triangleq \lim_{t\to\infty} \xi^i(t) \in \mathbb{R}^d, \quad \forall i \in \mathcal{V}$

Из A1 $\wedge$ A2 $\wedge$ A3 следует:

- $\boldsymbol{\xi}(\infty)$ – равновесие (C)
- $\xi^i(t)$ либо сходится к ξ^*, либо перестает меняться через конечное время

Заключение

1. Исследованы некоторые из новых классов моделей динамики мнений с ограниченным доверием.
2. Получены результаты о сходимости мнений для (A), (B) и (C).
3. При дополнительных предположениях для (A) установлена сходимость мнений за конечное время.
4. Дается классификация положений равновесия систем (A) и (B) и условия устойчивости.
5. Дополнительные предположения в теоремах существенны.