О комбинаторном анализе локальной полноты размеченных графов

Торшин И.Ю., Рудаков К.В.

ИОИ-2014

Содержание доклада

- Установление изоморфизма графов, полнота инвариантов графа
- О **комбинаторной теории разрешимости**
- Графы специального вида – **размеченные** графы
- Разметки χ-графов
- Инварианты на основе разметок χ-графов
- Критерии **локальной полноты** кортеж-инвариантов
- Комбинаторные оценки локальной полноты инвариантов
- Практические приложения

Установление изоморфизма графов, полнота инвариантов графа

- Граф - элемент множества $\boldsymbol{\Gamma} = \{(V, E) \mid V \subset N, E \subset N^2\}$
  «множество всех графов»
  N - натуральный ряд.
- Графы G_1, G_2 *тождественны* ($G_1 = G_2$), если $V_1 = V_2$ и $E_1 = E_2$
- Инцидентность вершины *v* и ребра *e*

$$v \bullet e \equiv (e = (v_1, v_2), (v = v_1) \vee (v = v_2))$$

- Графы G_1, G_2 *изоморфны* ($G_1 \simeq G_2$) - существует взаимно-однозначное соответствие между их вершинами и рёбрами, сохраняющее смежность вершин и инцидентность рёбер.

Th. $G_1 \simeq G_2 \Leftrightarrow \exists \mu_I : V_1 \to V_2 \mid G_2 = (\{\mu_I(v), v \in V_1\}, \{(\mu_I(u), \mu_I(v)), (u, v) \in E_1\})$

Инварианты графов

- *Инвариант* графа — числовая характеристика графа или упорядоченный список таких характеристик (кортеж), значение которой одинаково для каждого элемента произвольного класса изоморфных графов.

  $\iota : \mathbf{\Gamma} \to \mathrm{R}^{\mathrm{n}}, \mathrm{n} \in \mathrm{N}$ $\qquad \forall \mathrm{a} \in \mathbf{\Gamma} : \mathrm{b} \in \mathbf{I}(\mathrm{a}) \Rightarrow \iota(\mathrm{b}) = \iota(\mathrm{a})$
- *Элементарный инвариант* $\iota : \mathbf{\Gamma} \to \mathrm{R}$
- *Кортеж-инвариант* $\iota : \mathbf{\Gamma} \to \mathrm{R}^{\mathrm{n}}, \mathrm{n} \geq 2$
- В теории графов - десятки элементарных инвариантов:
  - сумма длин минимальных цепей между каждой парой вершин (индекс Винера)
  - кликовое число
  - число компонент связности графа
  - ….
- ***Условие полноты инварианта:*** $\forall \mathrm{a} \in \mathbf{\Gamma} : \mathrm{b} \in \mathbf{I}(\mathrm{G}) \Leftrightarrow \iota(\mathrm{a}) = \iota(\mathrm{b})$

О «локальной полноте» инвариантов

- Утверждения теории графов – по отношению к бесконечному множеству **Г**
- *Локальные формы утверждений о графах* формулируются по отношению к определенному *конечному* $\mathrm{P} \subset \Gamma$
- ***Целесообразность:*** *возможность комбинаторного тестирования выполнимости утверждений над* $\mathrm{P} \subset \Gamma$

В рамках комбинаторной теории разрешимости, графы рассматриваются как объекты, а их кортеж-инварианты - как вектора признаковых описаний объектов.

Основные положения комбинаторной теории разрешимости задач распознавания/классификации (1)

I. Определены *множество признаковых описаний и классы объектов*.

II. Задано конечное множество объектов («множество прецедентов»), описываемое совокупностью матрицы информации (признаки объектов) и информационной матрицы (принадлежность объектов к классам).

III. Множество объектов непротиворечиво, т.е. для произвольного набора признаков объекта существует только один ответ в информационной матрице.

Основные положения комбинаторной теории разрешимости (2)

IV. Непротиворечивость *множества объектов является необходимым условием разрешимости* (т.е. существования отображения из множества матриц информации во множество информационных матриц).

V. Разрешимость задачи при заданном множестве прецедентов может быть достигнута на определенных *подмножествах* множества признаков.

VI. При заданном подмножестве признаков, выполнимость критерия разрешимости устанавливается путем комбинаторного тестирования на исследуемом множестве прецедентов.

VII. Подмножества признаков могут выбираться на основании «информативности»

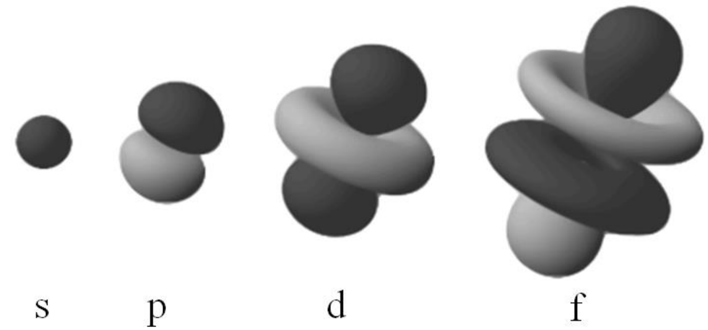
Поставленные задачи

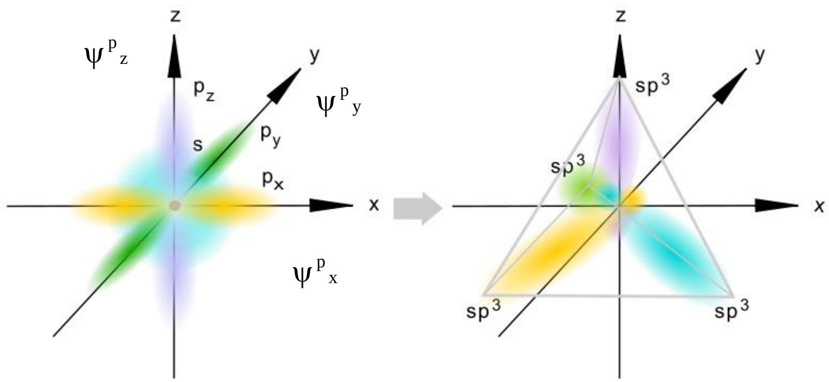
- Теоретическая задача
  – Разработка проблемно-ориентированной теории для анализа локальной полноты инвариантов *размеченных* графов
- Практические приложения
  – Критерии схожести хемографов (χ-графов)
  – Оценки скалярных молекулярных свойств (константы связывания и т.д.)
  – Анализ схожести графов состояний систем
  – Анализ семантических сетей
  – Анализ других размеченных графов (мультиграфов)

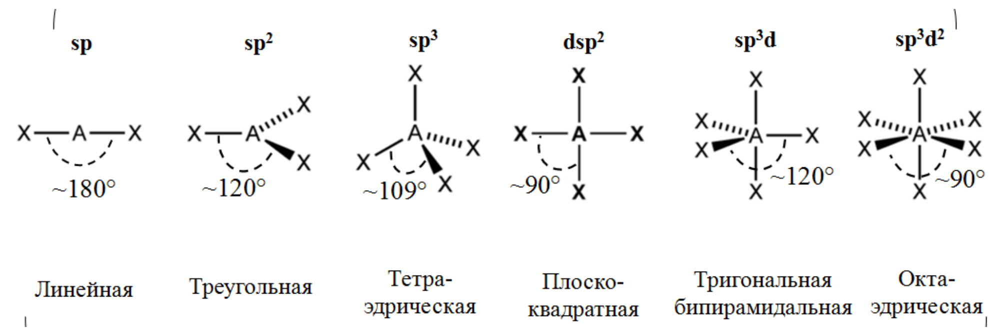
Графы специального вида – χ-графы

$$-\frac{\hbar^2}{2m}\Delta\psi(\vec{r})+U(\vec{r})\psi(\vec{r})=E\psi(\vec{r})$$

$$c_i(\psi_i^s+\sqrt{3}\psi_i^p)$$







Аксиома кратности химических связей $d+w_{max}=v_A+1$.

d - число связей атома
w_{max} – макс. кратность связи
v_A - макс. валентность

- **Определение.** *Хемографом X будем называть* ***конечный,*** *связный, неориентированный,* ***размеченный*** *граф без петель , с кликовым числом не превышающим 3 и для которого выполнена аксиома кратности.*

Аксиома кратности в терминах теории графов

$$d+w_{max}=v_A+1.$$

число смежных вершин в данном графе

максимальный элемент в соответствующей строке матрицы смежности данного графа

максимально возможное число ребер, инцидентных любой вершине в любом графе, соответствующей «типу атома» (т.е. метке вершины)

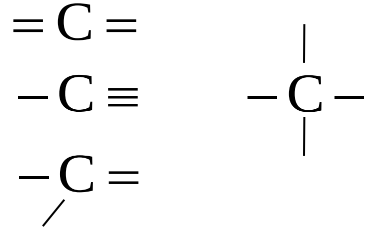
d - «число связей атома»

w_{max} – «макс. кратность связи»

v_A - «макс. валентность»

- Аксиома кратности:
  - Общее число ребер вершины с заданной меткой постоянно, а число инцидентных вершин - может изменяться.

Пример: метка «С»



Ограничения на словарь разметки и подмножества возможных сочетаний меток

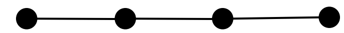
Разметки χ-графов

множество меток Y
функция разметки вершин $\mu_V : V \rightarrow Y$.

- В качестве меток можно использовать
  - химические типы атомов («C», «N», «O» и т. д.)
  - гибридные состояния атомов
  - максимальную валентность атома
  - заряды атомов
  - *комбинированные метки*
  - …

Аксиома кратности связей

χ-цепи



- Для порождения признаковых описаний хемографов вводятся понятия χ-цепей и χ-узлов.
  - X – размеченный граф, *множество меток* $Y=\{\upsilon_1, \upsilon_2, \ldots, \upsilon_{n(Y)}\}$, *функция разметки* $\mu_V : V \to Y$.
  - множество всех перестановок над Y - $\ddot{Y} = \bigcup_{n=1}^{\infty} Y^n$

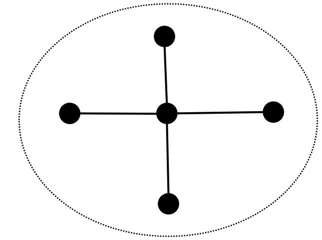
$$y^1 = (y_1^1, y_2^1, \ldots, y_i^1 \ldots, y_n^1) \quad y^1, y^2 \in \ddot{Y}$$
$$y^2 = (y_1^2, y_2^2, \ldots, y_i^2 \ldots, y_n^2)$$

  - ***χ-цепь*** α - элемент *множества χ- цепей,* $\alpha \in \tilde{Y}$

$$\tilde{Y} = \{\{y^1, y^2\}, y^1, y^2 \in \ddot{Y} \,\|\, |y^1| = |y^2| = n, \forall i = 1..n : y_i^1 = y_{n-i+1}^2\}$$

  - *Множество всех χ-цепей длины n* $\tilde{Y}^n \subset \tilde{Y}$

к- и χ-узлы



- *k-элементными сочетаниями* над Y - элементы множества $\sigma Y^{k} = \{\{y_1, y_2, \ldots, y_i \ldots, y_k\}, \mathop{\forall}\limits_{i=1}^{k} y_i \in Y\}$
- *χ-k-узел* к - элемент множества $\hat{Y}(k) = \{Y \times \sigma Y^{k}\}$
- ***χ-узел*** - элемент *множества* $\kappa \in \hat{Y} = \bigcup\limits_{k=2}^{8} \hat{Y}(k)$
- *К-узлы графа* - связные подграфы $(\Gamma(v), \hat{e}v)$
  - *множество смежности* вершины *v* $\Gamma(v) = \hat{v}\hat{e}v$
  - Все узлы χ-графа $\mathbf{K}(X) = \{(\Gamma(\mathbf{v}), \hat{\mathbf{e}}\mathbf{v}) \mid \mathbf{v} \in V(X),\ d(v) > 1\}$

$\mathbf{G} = (\mathbf{N}, \mathbf{N}^2)$ - бесконечный полный граф

$\mathbf{C} = \mathbf{C}(\mathbf{G})$ - множество всех цепей

$\mathbf{K} = \mathbf{K}(\mathbf{G})$ - множество всех узлов

Образующие множества подграфов

Объединение множества подграфов П $\quad \breve{\Pi} = \bigcup_{i=1}^{|\Pi|} \pi_i$

Множество всех замкнутых подграфов графа X(V,E)

$$\mathbf{\Pi}(X) = \left\{ (\mathbf{v}, \mathbf{e}) \;\; | \, \mathbf{v} \subseteq V, \mathbf{e} \subseteq E, \overset{\mathbf{e}}{\forall} (v_1, v_2) : v_1 \in \mathbf{v}, v_2 \in \mathbf{v} \right\}$$

Множество подграфов $O \subseteq \mathbf{\Pi}(X)$ ***- образует*** X, если $\breve{O} = X$

(покрытие всех вершин и всех рёбер графа X)

Множество всех цепей хемографа X

$$\mathbf{C}(X) = \left\{ c = (V_c, E_c) \,|\, c \in \mathbf{\Pi}(X), |\, E_c \,| = |\, V_c \,| - 1 > 0, \overset{|Vc|}{\underset{i=1}{\forall}} d(c, v_i) \le 2, \overset{|Vc|-1}{\underset{i=1}{\forall}} \; \overset{|Vc|}{\underset{j=i+1}{\forall}} v_i \ne v_j \right\}$$

Теорема 6. Множество всех цепей над χ-графом X образует X.

$$\breve{\mathbf{C}}(X) = X$$

$$\exists \mathbf{c} \subseteq \mathbf{C}(X) : \breve{\mathbf{c}} = X$$

Теорема 10. Множество всех узлов χ-графа X образует X.

$$\breve{\mathbf{K}}(X) = X$$

$$\exists \mathbf{\kappa} \subseteq \mathbf{K}(X) : \breve{\mathbf{\kappa}} = X$$

Отображения графов во множества разметок

- Теорема 11.$(\exists\mu_c : \mathbf{C} \to \tilde{Y}) \wedge (\exists\mu_\kappa : \mathbf{K} \to \hat{Y})$, т.е. множество цепей произвольного χ-графа X однозначно отображается в подмножество $\tilde{Y}(X)$ множества χ-цепей $\tilde{Y}$, а множество всех k-узлов X – в подмножество $\hat{Y}(X)$ множества χ-узлов $\hat{Y}$
  – Иначе говоря, существование функции разметки μ_V обуславливает существование функций $\mu_\kappa : \mathbf{K} \to \hat{Y}$ и $\mu_c : \mathbf{C} \to \tilde{Y}$
  – Следствие $\exists\mu_c^{-1} \Rightarrow \bigcup_{v \in V(X)} \hat{c}v = \bigcup_{\alpha \in \tilde{Y}(X)} \mu_c^{-1}(\alpha) = X$
  – Следствие $\exists\mu_\kappa^{-1} \Rightarrow \bigcup_{\kappa \in \hat{Y}(X)} \mu_k^{-1}(\kappa) = X$

- Существование обратных отображений μ_c^{-1} μ_κ^{-1} - необходимо для установления изоморфизма хемографов посредством установления полноты их инвариантов

Инварианты на основе разметок χ-графов

- Будем говорить, что *χ-цепь* $\alpha \in \tilde{\mathrm{Y}}$ *входит в хемограф* X, $\alpha \,\overline{\in}\, \mathrm{X}$, если $\alpha \in \tilde{\mathrm{Y}}(\mathrm{X})$.
- Соответственно, *вхождение χ-узла* $\kappa \in \hat{\mathrm{Y}}$ *в* X ($\kappa \,\overline{\in}\, \mathrm{X}$) соответствует $\kappa \in \hat{\mathrm{Y}}(\mathrm{X})$.
- ***Теорема 13.*** *Необходимые условия изоморфизма двух хемографов*

$$\tilde{\mathrm{Y}}(\mathrm{X}_1) = \tilde{\mathrm{Y}}(\mathrm{X}_2) \qquad \hat{\mathrm{Y}}(\mathrm{X}_1) = \hat{\mathrm{Y}}(\mathrm{X}_2)$$

  – χ-инварианты графа X –
    - *бинарные* $(\alpha \,\overline{\in}\, \mathrm{X})$ $(\kappa \,\overline{\in}\, \mathrm{X})$
    - *численные* $|\{\mathrm{x} \in \mathbf{\Pi}(\mathrm{X}) \,|\, \alpha \,\overline{\in}\, \mathrm{x}\}|$ $|\{\mathrm{x} \in \mathbf{\Pi}(\mathrm{X}) \,|\, \kappa \,\overline{\in}\, \mathrm{x}\}|$

Критерии полноты кортеж-*χ*-инвариантов

- *χ-инвариантами* будем называть инварианты хемографов, основанные на отношениях вхождения χ-цепей и χ-узлов в хемографы.
- *Оператор формирования χ-кортеж-инварианта* $\hat{\iota}\boldsymbol{\iota}_{\mathbf{e}}$ по множеству элементарных инвариантов $\boldsymbol{\iota}_{\mathbf{e}}$ - $\hat{\iota}\boldsymbol{\iota}_{\mathbf{e}} = (\iota_j, \iota_k, ..., \iota_l), \; \iota_j, \iota_k, ..., \iota_l \in \boldsymbol{\iota}_{\mathbf{e}}$
  - Значение i-го элемента кортежа $\hat{\iota}[i]\boldsymbol{\iota}_{\mathbf{e}}(G) = \iota(G) \mid \underbrace{\lambda(\iota) = i}_{\text{Функция нумерации элементарных инвариантов}}$
- *условие полноты инварианта*

  $$\forall a \in \boldsymbol{\Gamma} : b \in \mathbf{I}(G) \Leftrightarrow \iota(a) = \iota(b)$$

- ***Теорема 14.*** *Кортеж-инвариант является полным инвариантом тогда и только тогда, когда для произвольной пары неизоморфных графов в соответствующих кортежах значений данного инварианта существует различающийся элемент.*
  - *критерий полноты кортеж-инвариантов*

$$\forall a, b \in \boldsymbol{\Gamma} : \mathbf{I}(a) \cap \mathbf{I}(b) = \varnothing \Leftrightarrow \underset{i=1..n}{\exists} i : \hat{\iota}[i]\boldsymbol{\iota}_{\mathbf{e}}(a) \neq \hat{\iota}[i]\boldsymbol{\iota}_{\mathbf{e}}(b)$$

Комбинаторные оценки локальной полноты инвариантов

- ***Теорема 16.*** *Пусть задано множество прецедентов графов* Pr*, (* $\mathrm{Pr} \subset \mathbf{\Gamma}$*), множество χ-инвариантов* **χ** *и кортеж-инвариант* $\iota = \hat{\iota}\chi$ *. Тогда следующие утверждения эквивалентны:*
  - ι - *локально полный (над* Pr*),*
  - **χ** *обеспечивает разрешимость соответствующей задачи над Pr при непротиворечивых метках изоморфности,*

$$\forall_{a,b\in \mathrm{Pr}} \mathbf{iso}(a) \neq \mathbf{iso}(b) \Leftrightarrow \exists_{i=1..|\boldsymbol{\chi}|} i : \hat{\iota}[i]\chi(a) \neq \hat{\iota}[i]\chi(b)$$

$$\forall_{a,b\in \mathrm{Pr}} \mathbf{iso}(a) \neq \mathbf{iso}(b) \Leftrightarrow \exists_{i=1..|\boldsymbol{\pi}'|} i : \hat{\iota}[i]\hat{\boldsymbol{\beta}}[a]\boldsymbol{\pi}' \neq \hat{\iota}[i]\hat{\boldsymbol{\beta}}[b]\boldsymbol{\pi}'$$

Для инвариантов на основе цепей

  - *над регулярным Pr,* **χ** *обеспечивает регулярность*
  - *Задача распознавания изоморфных графов разрешима тогда и только тогда, когда для каждого графа из Pr разность класса изомерных по* ι *графов и класса изоморфных графов пуста.*

$$\forall_{a\in \mathrm{Pr}} \mathbf{i\mu}(a, \iota, \mathrm{Pr}) \setminus \mathbf{i}(a, \mathrm{Pr}) = \varnothing$$

    - *локальное множество изоморфных графов* $\mathbf{i}(G, \mathrm{Pr}) = \{g \in \mathrm{Pr} \mid iso(g) = iso(G)\}$
    - *локальное множество изомерных графов* $\mathbf{i\mu}(G, \iota, \mathrm{Pr}) = \{g \in \mathrm{Pr} \mid \iota(g) = \iota(G)\}$
  - *Пусть* $r_\iota(\iota, \mathrm{Pr}) = 1 - \frac{1}{|\mathrm{Pr}|^2} \sum_{a\in \mathrm{Pr}} |\mathbf{i\mu}(a, \iota, \mathrm{Pr}) \setminus \mathbf{i}(a, \mathrm{Pr})|$*. Тогда инвариант полон тогда и только тогда, когда* $r_\iota(\iota, \mathrm{Pr}) = 1$ *.*

Комбинаторная оценка локальной полноты кортеж-инварианта

Пример практического приложения

- Функции расстояния между хемографами
  - *метрика Хэмминга* над бинарными χ-инвариантами

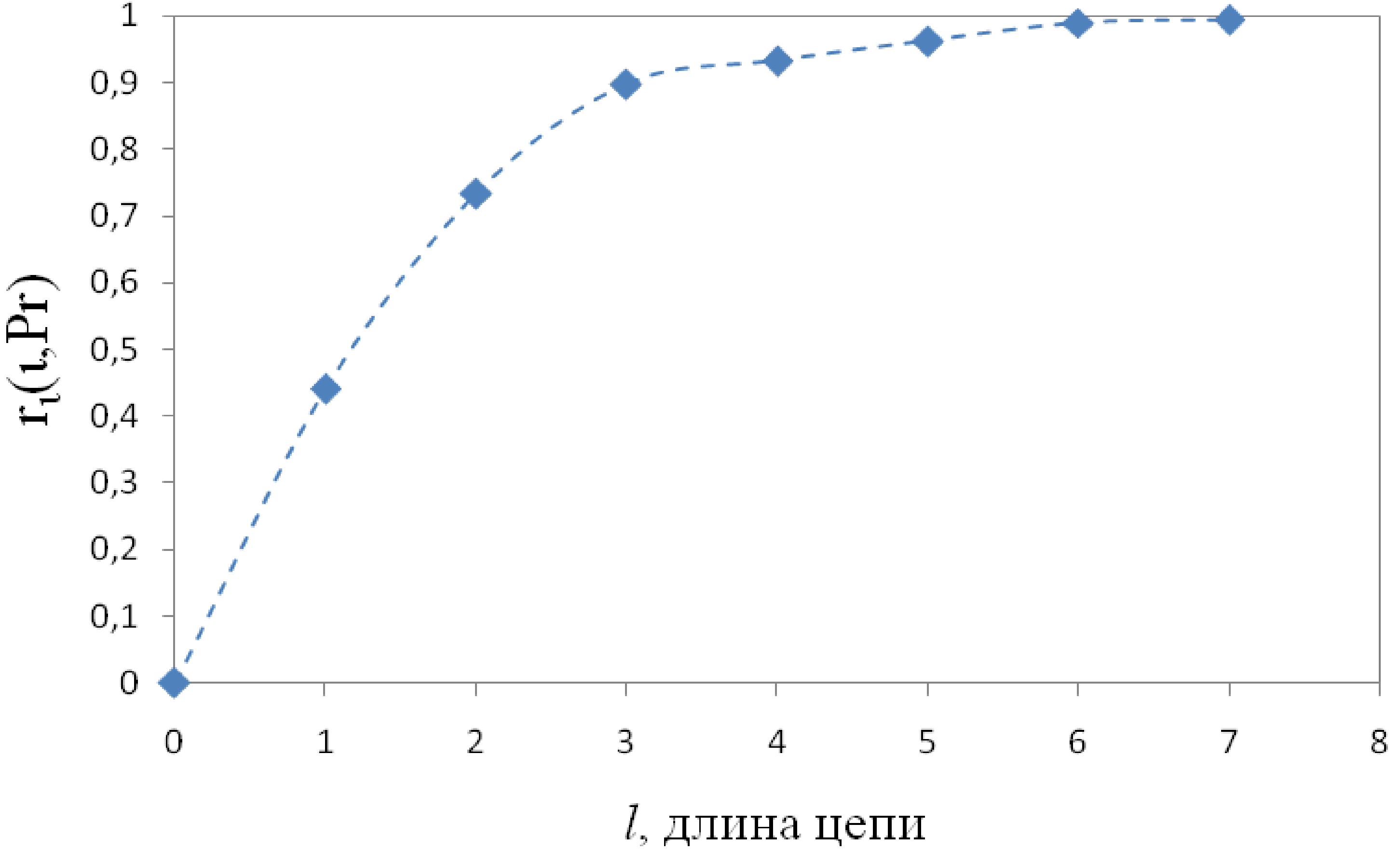
$$d_{\chi b}(X_1, X_2) = \frac{1}{n}\sum_{i=1}^{n} \hat{\imath}[i]\hat{\boldsymbol{\beta}}[X_1]\boldsymbol{\pi} \oplus \hat{\imath}[i]\hat{\boldsymbol{\beta}}[X_2]\boldsymbol{\pi}$$

  - *метрики Минковского* над множеством численных χ-инвариантов

$$d_{\chi\eta}(X_1, X_2) = \sqrt[\frac{1}{p}]{\sum_{i=1}^{n} \left|\hat{\imath}[i]\hat{\boldsymbol{\eta}}[X_1]\boldsymbol{\pi} - \hat{\imath}[i]\hat{\boldsymbol{\eta}}[X_2]\boldsymbol{\pi}\right|^p}$$

- Тестирование
  - множество меток Y={H, C_{sp3}, C_{sp2}, C_{sp1}, N_{sp3}, N_{sp2}, N_{sp1}, O_{sp3}, O_{sp2}, P_{sp3}, S_{sp3}, X_{sp}, X_{sp2}, A}
  - 500000 структур PUBCHEM
  - семейства кортеж-инвариантов (n=1..7, k=3, 4)
    - бинарные инварианты над множеством χ-цепей
    - бинарные инварианты над множеством χ-узлов из k вершин

Расчет локальной полноты ($r_\iota(\iota, Pr)$) кортеж-инвариантов над χ-цепями фиксированной длины

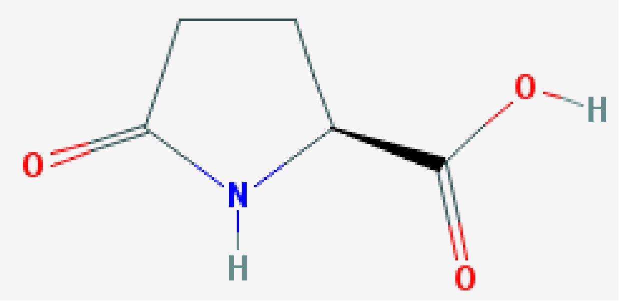


Y^n n=1..7

- n=7 оценка достигла значения 0.99±0.01.
- при сравнительно коротких длинах χ-цепей (n=5) - 0.96 ±0.02.

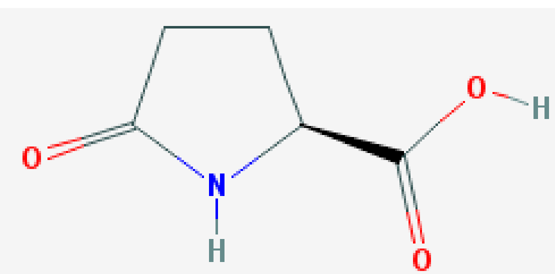
Поиск схожих молекул

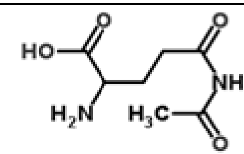
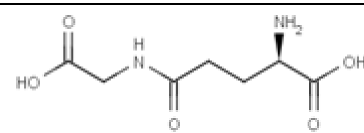
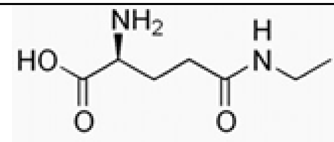
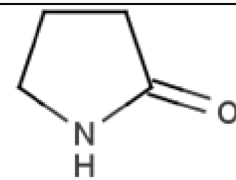
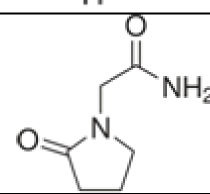
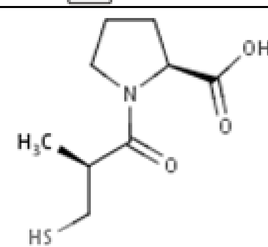
- L-пироглутаминовая кислота - эффективный переносчик магния внутрь клеток.



- В клинической практике, данное соединение характеризуется выраженными неврологическими эффектами.
- Долгое время стоял вопрос о том, каким образом эти эффекты могут осуществляться на уровне молекулярных механизмов.

Соединения, химическая структура которых схожа с пироглутаматом и которые объясняют нейротропные свойства пироглутамата



$d_{\chi b}$	Структурная формула	Название вещества	Медико-физиологическое значение
0.07		N-ацетилглутамин	Детоксикация азотистых оснований
0.22		5-глутамилглицин	Антагонист глутамата (нейропротекция)
0.37		L-теанин	Антагонист глутамата, ноотропный эффект
0.45		2-пирролидинон	Антагонист глутамата, активация ацетилхолиновых рецепторов
0.45		Пирацетам	Высвобождение ГАМК, секреция ацетилхолина, антагонист глутамата, ноотропные эффекты
0.50		Каптоприл	Ингибитор АПФ, нормализация АД

Выводы

- Формализм, позволяющий применять комбинаторную теорию разрешимости к теоретико-графовым построениям.
- Особые виды разметок –χ-цепи и χ-узлы, отношения вхождения которых в размеченный граф являются инвариантами.
- Получены критерии локальной полноты исследуемых наборов инвариантов и проведено их комбинаторное тестирование
- Результаты практического применения предлагаемого формализма.