

## ПОШУК ІДЕНТИЧНОЇ СТРУКТУРИ В МОЛЕКУЛЯРНИХ ГРАФАХ ХІМІЧНИХ СПОЛУК

Остапенко О.О.

Науковий керівник – к.т.н., доц. Чала Л.Е.

Харківський національний університет радіоелектроніки  
(61166, Харків, просп. Науки, 14, каф. ШІ, тел. (057) 702-13-37)

This work is devoted to research on the relationship between the structure of a chemical compound and its properties. It scrutinizes the problems of finding identical structures in chemical compounds represented by molecular graphs. The measure of similarity between the two compounds is determined by metric estimates of proximity. It suggests an isomorphic subgraph search model, extended by combining metric estimates and vertex differentiation methods. The model can be used to improve searching of chemical databases.

Створення інтелектуальних систем, заснованих на багаторівневих онтологіях хімії, потребує ефективного представлення структур та властивостей хімічних сполук за допомогою адекватних математичних моделей. Доведено, що властивості сполук є в першу чергу наслідком схеми зв'язку атомів в молекулі [1]. Тому при моделюванні зв'язків між структурою хімічної сполуки та її властивостями важливе значення набуває проблема ідентичності.

До основних напрямків визначення ступеня структурної близькості двох хімічних сполук  $G$  та  $H$  відносять:

– методи знаходження перетину  $G \cap H$  – максимальної загальної частини хімічної структури для молекул  $G$  и  $H$  (алгоритми пошуку MCS – Maximum Common Subgraph);

– методи представлення структур  $G$  та  $H$  у вигляді множин молекулярних фрагментів малого розміру, ступень близькості визначається як функція числа співпадаючих фрагментів та кількості фрагментів, що відрізняються.

Для внутрішнього представлення структур хімічних сполук використовуються молекулярні графи. Кожному фрагменту молекулярної структури відповідає певний підграф молекулярного графа. Залежність «структура-властивість» базується на гіпотезі, що загальні властивості деякого класу обумовлені наявністю загального структурного фрагменту в кожному з його членів. Такий фрагмент може складатися із декількох не пов'язаних між собою частин. Властивості з'єднання можуть залежати як від положення загального фрагмента всередині структури, так і взаємного положення непов'язаних частин загального фрагмента. Тобто потрібно знайти істотні фрагменти, які визначають властивості даної групи сполук, що дозволить прогнозувати властивості сполук, які включають ці фрагменти.

Розглянемо цю задачу в термінах теорії графів. Граф  $F$  є спільним підграфом графів  $G$  та  $H$ , якщо в них існують підграфи  $G'$  та  $H'$ , які ізоморфні

графу F. Спільний підграф двох графів є найбільшим, якщо цей підграф містить максимальну кількість вершин та/або ребер серед інших спільних підграфів. Перетин графів G та H – найбільший по кількості вершин спільний підграф цих графів, а перекриття графів G та H – найбільша по кількості ребер частина. Тобто пошук подібної хімічної сполуки зводиться до визначення найбільшого ізоморфного підграфа. На його основі, використовуючи функції відстаней [2], можна оцінювати міру близькості двох сполук.

Розглянемо процедуру пошуку ізоморфної структури на множині неорієнтованих молекулярних графів  $Q^{(n)}$ . Нехай  $Q^\cap$  – множина їх найбільших спільних підграфів,  $Q^\cap = \{Q_{G,H} \mid \forall G, H \in Q^{(n)}, Q_{G,H} = Q_{H,G}, G \neq H\}$ .

Тоді модель пошуку ізоморфного підграфу можна представити у вигляді:

$$G \cap H \Rightarrow \tilde{S}(G) \cap \tilde{S}(H), \tilde{M}^2(G) \cap \tilde{M}^2(H) \Rightarrow Q_{G,H} \{(v_{g,h}, u_{g,h})\} \Rightarrow r_{g,h}, R'_{g,h},$$

де  $S(Q)$  – ступеневий інваріант графа Q, розраховується для кожної вершини графа як число ребер, інцидентних даній вершині;

$\tilde{S}(Q)$  – упорядкована ступенева характеристика, вектор упорядкованих за зростанням ступеня вершин графа;

$M^2(Q)$  – маршрутний інваріант графа, відображує для кожної вершини відповідну упорядковану строку матриці  $A^2(Q)$ ,  $a_{ij} \neq 0$ ;

$\tilde{M}^2(Q)$  – упорядкований маршрутний інваріант графа, вектор упорядкованих за зростанням маршрутних характеристик вершин графа;

$v_{g,h}, u_{g,h}$  – множина вершин та ребер найбільшого ізоморфного підграфа;

$r_{g,h}, R'_{g,h}$  – абсолютні та відносні відстані, що відповідають умовам метрик.

Запропонована модель є розвитком метода знаходження максимальної спільної частини в структурі сполук. Пошук зводиться до диференціації вершин обох молекулярних графів та знаходження їх перетинів, що визначають найбільший ізоморфний підграф, та обчислення функцій відстаней, що служать мірою подоби для двох молекулярних графів.

#### Список литературы

1. Bunke H. Comparison of Algorithms for Maximum Common Subgraph on Randomly Connected Graphs [Text] / H. Bunke, C. Guidobaldi, P. Foggia, etc. // Structural, Syntactic, and Statistical Pattern Recognition, Joint IAPR International Workshops SSPR 2002 and SPR 2002, Windsor, Ontario, Canada, August 6–9, 2002. — P. 123–132.

2. Макаров Л. И. Метрические свойства функций расстояний между молекулярными графами [Текст] / Л. И. Макаров // Журнал структурной химии. — 2007. — Том 48, № 2. — С. 223 – 229.