

Быстрый алгоритм идентификации подобных химических соединений

А.П. Сергеев

ООО «Диалектика», проспект Академика Глушкова, д. 2, корп. 6, к. 214, Киев, Украина

a_p_sergeyev@mail.ru

Аннотация. Рассматривается быстрый алгоритм идентификации подобных химических соединений. Назначение алгоритма – поиск и фильтрация идентичных (с точностью до оптической и пространственной изомерии) органических химических соединений. Алгоритм может применяться для распознавания новых химических соединений, а также для поиска изомеров (подобных соединений) в онлайн-овых и оффлайн-овых химических базах данных. Выполнена оценка вычислительной сложности алгоритма.

Ключевые слова

Параллельные вычисления, молекулярный граф, изоморфизм, изоморфные деревья, перенумерация вершин.

1 Введение

Ежегодно в мире синтезируется более 250 тысяч химических соединений, которые хранятся в химических базах данных, как онлайн-овых (PubChem), так и оффлайн-овых (MAFR). Причем многие химические базы данных доступны на бесплатной основе. Но при работе с базами данных возникает ряд проблем, некоторые из которых перечислены ниже.

Определение новых соединений. После синтеза нового химического соединения определяется его структурная формула, а затем выполняется поиск в базах химических соединений в целях поиска подобных структурных формул. На этапе поиска быстрый алгоритм идентификации химических соединений позволит сэкономить время и вычислительные ресурсы.

Синтез соединений с наперед заданными свойствами. Часто возникает задача синтеза соединений, свойства которых аналогичны свойствам имеющегося соединения, но цена существенно меньше. Например, в фармакологии наравне с основным лекарством используются десятки, а то и сотни дженериков, которые стоят в разы меньше оригинального лекарства. В процессе синтеза дженериков можно воспользоваться быстрым алгоритмом идентификации подобных химических соединений позволит ускорить процесс отбора соединений с требуемыми свойствами.

2 Теоретическая часть

Согласно [3], подобные химические соединения обладают подобными свойствами. Структурные формулы химических соединений представляются *молекулярными графами* ([1], [2]), представляющие собой связанные неориентированные графы, вершины которых соответствуют атомам молекулы, а ребра — связям между атомами молекулы. Справедливы следующие утверждения.

Утверждение 1. Подобным химическим соединениям соответствуют изоморфные молекулярные графы.

Из справедливости этого утверждения следует возможность применения быстрого алгоритма фильтрации изоморфных XSD-схем ([5]). В рассматриваемом случае вместо XSD-схем на вход алгоритма поступают молекулярные графы.

Утверждение 2. *Химические соединения с изоморфными молекулярными графами являются изомерами (оптические и пространственные).*

В соответствии с этим утверждением можно воспользоваться алгоритмом [5] для поиска и фильтрации молекул изомеров химических соединений.

Утверждение 3. *Подобные химические соединения обладают сходной биологической активностью [4].*

Из утверждений 2 и 3 вытекает, что изомеры химических соединений могут также обладать похожей биологической активностью.

С помощью алгоритма [5] можно конструировать структурные молекулы подобных химических соединений. Благодаря этому открывается возможность целенаправленного создания дженериков, стоимость которых существенно меньше стоимости оригинальных лекарств.

Алгоритм идентификации химических соединений целесообразно использовать для поиска и идентификации соединений как в оффлайновых, так и в онлайн-базах данных. В последнем случае требуется разработка веб-интерфейса, обеспечивающего обмен данными между химической базой данных и алгоритмом.

Оценка вычислительной сложности. Как отмечалось в [5], в большинстве случаев вычислительная сложность применяемого алгоритма оценивается как $O(n^2)$, где n – количество вершин молекулярного графа. Благодаря распараллеливанию, применяемому по отношению к разбитым на классы потенциально подобным химическим соединениям, дополнительно повышается эффективность работы алгоритма. При этом максимальный прирост скорости вычислений составит $m/2$, где m – количество процессоров в многопроцессорной системе.

3 Выводы

Рассмотрено применение быстрого алгоритма фильтрации изоморфных XSD-схем для идентификации подобных химических соединений. Получена оценка вычислительной сложности алгоритма.

Литература

- [1] Р. Кинг. Химические приложения топологии и теории графов. *М.: Мир*, 1987..
- [2] Р. Бейдер. Атомы в молекулах. Квантовая теория. *М.: Мир*, 2001.
- [3] А.М. Johnson, G.M. Maggiora. Concepts and Applications of Molecular Similarity. New York: John Willey & Sons, 1990.
- [4] Н. Kubinyi (1998). «Similarity and Dissimilarity: A Medicinal Chemist's View». *Persp. Drug Discov. Design* 9-11: 225–252.
- [5] А.П. Сергеев. Быстрый алгоритм фильтрации изоморфных XSD-схем. *Материалы конференции HPC-UA'2011*, Киев, 2011, с. 122-126.

Fast algorithm of identification of conformable chemical compounds

Abstract. *It is considered fast algorithm of identification of conformable chemical compounds. Purpose of algorithm – search and a filtration identical (to within an optical and spatial isomerism) organic chemical compounds. The algorithm can be applied to recognition of new chemical compounds, and also to search of isomers (conformable compounds) in online and offline databases. The estimation of calculating of complication of such algorithm is gained.*