

令和 6 年 6 月 16 日現在

機関番号：34419

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K12000

研究課題名（和文）Deep Graph Kernelを用いた進化計算による構造化データの探索

研究課題名（英文）Evolutionary Search of Structured Data by Using Deep Graph Kernels

研究代表者

半田 久志（Handa, Hisashi）

近畿大学・情報学部・教授

研究者番号：60304333

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究課題では、特徴空間を構成するときに、特徴間の関係を埋め込むことより、カーネル関数の特性に左右されない安定した探索を実現する。事前学習が必要となるが、領域知識が提供できる問題クラスなら適用可能であると考えた。特に、応用分野として有機薄膜太陽光電池の分子構造探索で評価する。有機薄膜太陽電池は、大きい分子構造となるため、スクラッチで原子を組み合わせても解を探索することができない。そこで、既存の有機薄膜太陽電池に用いられている分子構造の部分構造をモジュールとして、モジュールの組合せを探索するようにした。また、モンテカルロツリーサーチを用いてモジュールの設計について考察し、それらの有用性を確認した。

研究成果の学術的意義や社会的意義

応用分野として最小した有機薄膜太陽電池は、地球温暖化対策として化石燃料に代わるクリーンなエネルギー資源として広く注目されている。主に使用されているシリコン半導体を用いた太陽電池は、半導体を使用するため製造コストが高く、さらに設置面積が広いため設置場所の自由度が低い。そのため利用範囲が限定されている。有機薄膜太陽電池はシリコン半導体の代わりに有機半導体を用いて製造される新たな太陽電池である。有機半導体は製造過程が簡便であるため安く大量生産することが可能で、さらに折り曲げに強く柔軟であるといった性質を持つことから、低コストかつ自由な場所に設置できる次世代の太陽電池として期待されている。」

研究成果の概要（英文）：In this research project, by embedding the relationship between features when constructing the feature space, a stable search that is not affected by the characteristics of the kernel function is achieved. Although prior learning is required, we considered that it is applicable to any problem class for which domain knowledge can be provided. In particular, we will evaluate it in the molecular structure search of organic thin-film photovoltaic cells as an application area. Since organic thin-film photovoltaic cells have a large molecular structure, it is not possible to search for a solution by combining atoms in a scratch. Therefore, we used substructures of molecular structures used in existing organic thin-film photovoltaic cells as modules to search for combinations of modules. We also discussed the design of modules using Monte Carlo tree search and confirmed their usefulness.

研究分野：進化計算

キーワード：Deep Graph Kernel 有機薄膜太陽電池 進化型多目的最適化 モンテカルロツリーサーチ

## 1. 研究開始当初の背景

本研究課題では、グラフの探索を進化計算にて行うものである。先行研究の **EDA-GK** では、特徴空間で問題を捉えることにより、遺伝子型から表現型への写像の凸凹さを解消できていると考えられる。ところが、研究を押し進めていく上で、先行研究の **EDA-GK** では、カーネル関数が対象としている特徴空間によって、問題によっては機能しないことが判ってきた。例えば、前ページの最小パス距離カーネルの場合、解が全結合に近いグラフであった場合、任意のノード間の距離がほとんど1となり、最適解付近のグラフのヒストグラムに違いがなくなる。結果、ランダムなエッジの追加・削除によるランダム探索に近い挙動を示すことになる。

そこで、本研究課題では、**Deep Graph Kernel** の利用を検討する。**Deep Graph Kernel** はグラフカーネルが対象とする特徴に対して、それらの関係性も学習するものである。**word2vec** にあるような単語の埋め込みをグラフに拡張したものである。特徴空間を構成するときに、特徴間の関係を埋め込むことより、カーネル関数の特性に左右されない安定した探索を実現する。事前学習が必要となるが、領域知識が提供できる問題クラスなら適用可能であると考え、特に、提案手法を有機薄膜太陽光電池で評価することを考える。

## 2. 研究の目的

有機薄膜太陽電池は、比較的大きめの分子構造となるため、スクラッチで原子を組み合わせても適切な解表現を探索することができない。そこで、既存の有膜太陽光電池に用いられている分子構造の部分構造をモジュールとして、モジュールの組み合わせを探索するようにした。さらに、**QDF** という深層学習の一種を用いることにより、より大規模な分子構造をもつ解に対して、進化計算が適用できるようになった。また、モンテカルロツリーサーチを用いてモジュールの設計について考察した。

## 3. 研究の方法

本研究では候補解となる化合物の構造を構成する際に、あらかじめデータベースに蓄積した化合物を組み合わせる手法を採用する。また、部分化合物を蓄積したデータベースを化合物リストと呼ぶ。化合物リストには、ラベル名、化学式、SMILES、結合可能な原子のリスト (joints) が CSV 形式で記されている。化合物リストに既存のアクセプター分子で使われている分子の部分構造の情報を追加する。以下にデータベースの一部を示す：

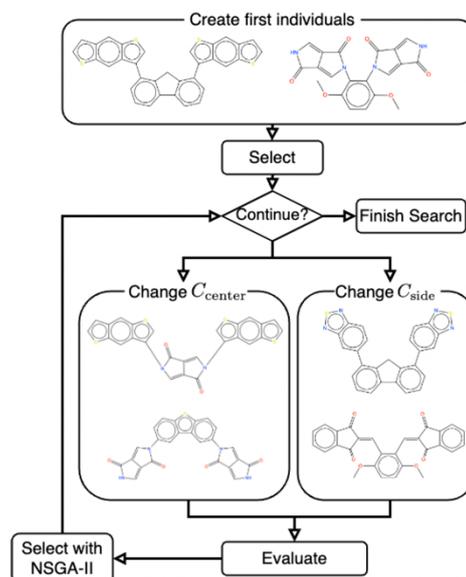
ラベル名	化学式	SMILES	joints
B1	$C_6H_6$	<chem>C1=CC=CC=C1</chem>	0:1:2:3:4:5
B2	$C_{10}H_8$	<chem>C1=CC=C2C=CC=CC2=C1</chem>	0:1:2:4:5:6:7:9
C1	$C_{13}H_{10}$	<chem>C1C2=CC=CC=C2C3=CC=CC=C31</chem>	0:2:3:4:5:8:9:10:11
C2	$C_{20}H_{14}$	<chem>C1C2=CC=CC=C2C3=C1C4=C(C=C3)C5=CC=CC=C5C4</chem>	0:2:3:4:5:11:12:14:15:16:17:19
C3	$C_{13}H_{11}N$	<chem>CN1C2=CC=CC=C2C3=CC=CC=C31</chem>	0:3:4:5:6:9:10:11:12
D1	$C_{12}H_7NO_2$	<chem>C1=CC2=C3C(=C1)C(=O)NC(=O)C3=CC=C2</chem>	0:1:5:8:12:13:14
D2	$C_6H_4N_2O_2$	<chem>O=C1NC=C2C(=O)NC=C12</chem>	0:2:3:7:8
E1	$C_4H_3NOS_2$	<chem>C1C(=O)NC(=S)S1</chem>	0:4
E2	$C_{10}H_4O_2S_2$	<chem>C1=C2C(=CS1)C(=O)C3=CSC=C3C2=O</chem>	0:3:8:10

本研究では化合物リストに蓄積された部分化合物を組み合わせることで、新たな化合物を構成する。その際に重要となるのは組み合わせる方法だが、合成の容易さを考慮するために対称な構造を持つものが望ましい。そのため本研究では部分化合物  $C_{center}$  の左右に、部分化合物  $C_{side}$  を配置した構造に限定して探索を行う。

本研究では、解を連続的な空間に配置せず、目的関数が微分可能でないため、進化計算の手法に従って、ランダムな初期値設定、子世代の生成、選択を行う。探索プロセスの概要図は次ページ右上の図に示す。この探索プロセスでは、**QDF** の指標に基づいて探索を行っているため、得られた解の候補の性能は実験や量子化学シミュレーションによって検証する必要がある。子世代の生成においては、形質を維持するために、 $C_{center}$  と  $C_{side}$  を同時に変更するのではなく、一方のみを変更する。この手順は、世代ごとに交互に行われる。子個体生成時に **Deep Graph Kernel** を用いてモジュールからの部分化合物の選択を確率的に行った。選択プロセスでは、優れた個体

が順に維持される truncation 戦略, 多様な最適解の集合を生成する NSGA-II の 2 つを採用した。また, 評価プロセスでは, QDF を使用して化合物の振動子強度と波長からスペクトルを積分した値を予測している。さらに RD-Kit で用いられている合成のしやすさである SAScore に基づいた評価関数も導入した進化型多目的最適化についても調査した。

また, モンテカルロツリーサーチにより, パーツなる部分化合物の探索を行った。モンテカルロツリーサーチ [1] は, 有望なノードのサブツリーを優先的に探索することにより, 広大な探索空間を効率的に探索する探索手法である。初めはルートノードのみが存在し, 探索を進めるにつれ徐々にツリーが構築される。モンテカルロツリーサーチの大きな特徴は, ツリー自体は浅いレベルで構築され, リーフノードへの完全なパスはランダムなプレイアウトにより生成されるところにある。これによって, 広大な探索空間での効率的な探索を実現する。プレイアウトによって解が得られると, ルートノードへのパスが通るすべてのノードに評価スコアが伝達される。モンテカルロツリーサーチにより生成された分子を, 有機薄膜太陽電池の材料として優秀な既存の化合物の一部分と置換する形で合成する。置換対象の分子は, 既存化合物の Scaffold 生成時に取り除かれた分子のうち 対称性をなす一組から選択する。図 5 は, 化合物 Y6 において置換対象となる部分を表した図で, それぞれ同.じ番号でマークされた部分同士が対称な組となっている。

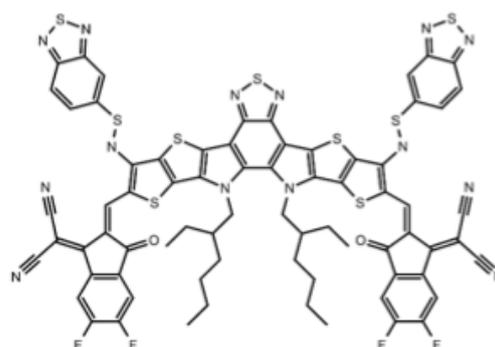
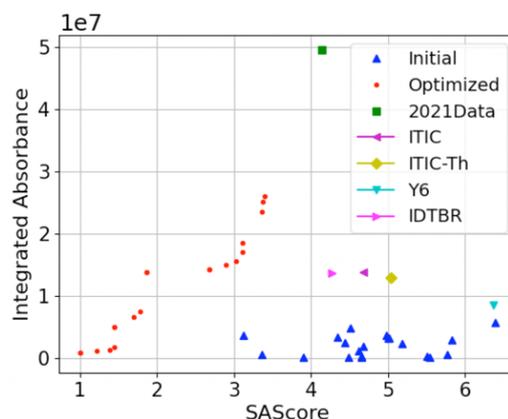


#### 4. 研究成果

進化型多目的最適化による探索結果を右のグラフに示す。SAScore (合成のしやすさ) は最小化関数で, 光吸収量は最大化の関数であるため, 左上にいくほど良い解となる。緑の点は再帰的に提案手法を適用したものであり, 分子サイズが大きいものである。赤の点が今回の提案手法による結果で, 分子サイズが小さいこともあり, 全体的な吸収量は減ってしまっている。分子サイズで正規化した関数を用いれば良いのではないかと考えている。ITIC や Y6 は既存の有機薄膜太陽電池であり, QDF を用いた光吸収量では提案手法による改善が見られている。しかしながら, Gaussian を用いた場合は, そこまで差が顕著ではないので, QDF の精度改善が必要である。

Y6 の評価スコアを上回るスコアとなっていることがわかるため, QDF が予測する光吸収性能において Y6 よりも優秀な化合物を発見することができたといえる。右下の図に提案された化合物を示す。モンテカルロツリーサーチの探索結果に対して, Y6 の一部分を変えることで新たな化合物を生成している。

提案した化合物のスペクトルが既存の化合物よりも全体的に高波長側に寄っていることがわかった。太陽光は波長 500nm 付近で最も強い光を放つため, この波長でより高い光吸収性能となることが望ましいが, 一部の化合物では達成できていないことがわかった。原因として考えられるのは, 評価スコアの算出方法である。現状の算出方法では太陽光のスペクトル分布を考慮していないため, 各波長における評価の重み付けされていない。これを解決する方法として, 化合物のスペクトルと太陽光のスペクトルの積を評価スコアとして利用することが考えられる。



5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計5件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 木原泰一, 山田武士, 半田久志, 大久保貴志
2. 発表標題 有機薄膜太陽電池のためのモンテカルロツリーサーチを用いた分子設計
3. 学会等名 計測自動制御学会システム・情報部門学術講演会2022
4. 発表年 2022年

1. 発表者名 岩寄智佳良, 葉山大雅, 半田久志, 大久保貴志
2. 発表標題 有機薄膜太陽電池に適した光吸収量の多い化合物の進化的探索
3. 学会等名 FANシンポジウム2022
4. 発表年 2022年

1. 発表者名 葉山大雅, 高石優也, 半田久志
2. 発表標題 分子の特徴量と Deep Graph Kernels を用いた化合物の表現方法に関する一考察
3. 学会等名 電気学会 電子・情報・システム部門大会
4. 発表年 2021年

1. 発表者名 葉山大雅, 半田久志, 大久保貴志
2. 発表標題 QDF を用いた有機薄膜太陽電池の分子構造の進化的探索
3. 学会等名 計測自動制御学会 システム・情報部門 学術講演会 2021
4. 発表年 2021年

1. 発表者名 森川 大樹, 葉山 大雅, 半田 久志
2. 発表標題 有機薄膜太陽電池のアクセプター分子のための分子構造局所探索手法
3. 学会等名 第63回システム工学部会研究会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関