

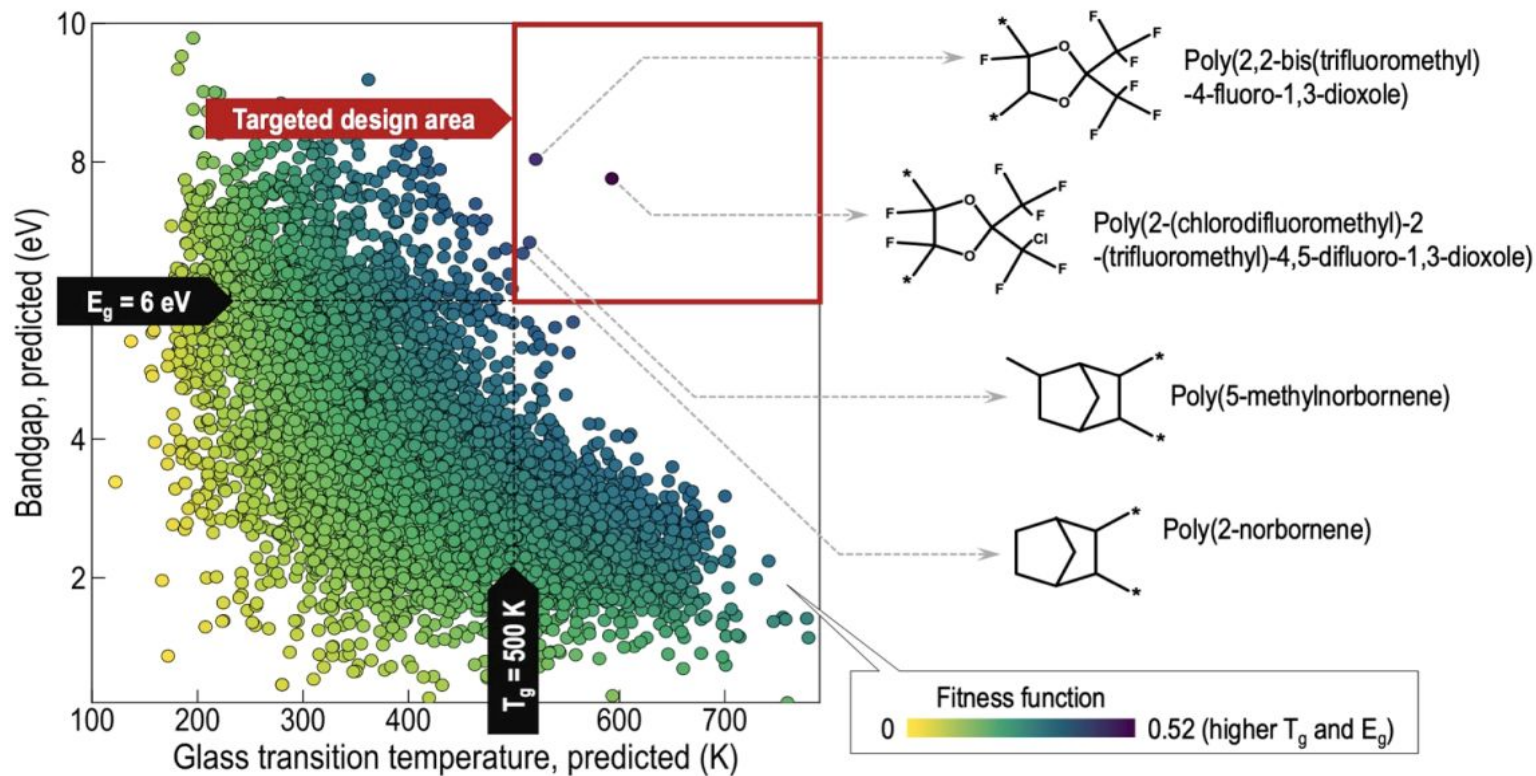
Polymer design using genetic algorithm and machine learning

Radovan Hodor, Michal Voronič
Technical University of Košice

Definícia problému a využitie EA

- Polyméry sú zložité periodické molekuly zložené z monomérov.
- Riešili **inverzný problém** - chceli aby polyméry mali isté vlastnosti.
- Na vytváranie nových chemických molekúl sa využívajú aj autoenkódery a GAN, ale na periodické molekuly to zatiaľ nie je efektívne.
- V tomto prípade chceli dve vlastnosti - vysokú tepelnú stabilitu a schopnosť zniesť vysoké elektrické pole.
- Dovtedy sa vedelo o **4 z cca 12000** popísaných polyméroch s týmito vlastnosťami.

Definícia problému a využitie EA



Problém definovaný ako EA

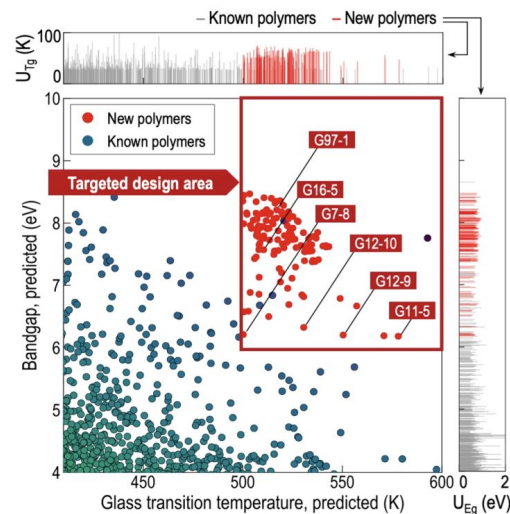
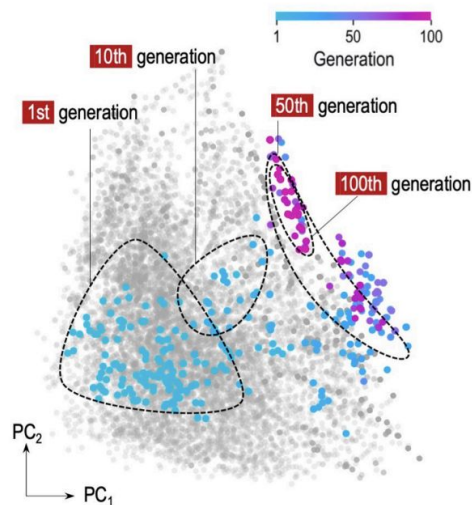
- Jedinec je definovaný ako **sekvencia základných chemických blokov** spojených kovalentnými väzbami.
- Na týchto jedincoch sa jednoducho robia **mutácie a kríženia**.
- Fitness funkcia bol spočiatku problém, lebo neexistuje žiadne presný výpočet na hľadané charakteristiky. Ale podarilo sa im vyvinúť jednoduchý **ML model, ktorý vie v krátkom čase predikovať ohodnotenie pre jedinca**. Hodnoty pre tepelnú stabilitu a schopnosť zniesť vysoké elektrické pole normalizovali, lebo tepelná stabilita bola stonásobne vyššia hodnota ($E_q \times T_q$).
- Z cca **12000** popísaných polymérov extrahovali **3045** základných chemických blokov.
- V prvej generácii bolo vytvorených 100 polymérov pozostávajúcich z 8 stavebných blokov. Stavebné bloky boli vybrané náhodne rešpektujúc ich frekvencie výskytu v referenčnom súbore údajov o polyméroch. V každej iterácii EA bolo **10 najlepších polymérov s najvyššou fitness** zachovaných, aby vytvorili potomkov ďalšej generácie prostredníctvom kríženia a mutácie.

Problém definovaný ako EA

- Počas **kríženia** dva rodičovské polyméry generujú potomka kombinovaním jedného náhodného segmentu prvého polyméru s iným náhodným segmentom druhého polyméru.
- **Segmentačný bod** základného polyméru bol zvolený pomocou **Gaussovej funkcie** so strednou hodnotou smerujúcou do stredu monomérnej jednotky (pri polyméroch s 8 stavebnými blokmi v monomérnej jednotke sa segmentácia **vyskytuje väčšinou v ich strede**, pričom každý výsledný polymérny segment obsahuje väčšinou 4 bloky)
- Umožnili malú variáciu výberu polohy segmentácie, aby sa monomérna jednotka mohla **rozdeliť na fragmenty s väčším alebo menším počtom blokov**. Pre každý rodičovský polymér sa **segmentačný proces uskutočnil 4-krát**, výsledkom čoho bolo celkovo 180 potomkov polymérov z 10 rodičovských polymérov v každej iterácii.
- Pridali aj mutáciu, ktorá bola náhodná zmena chemického bloku s 40% pravdepodobnosťou. Jednotlivcov, ktorí **porušili iné chemické pravidlá zahodili**.

Výsledky

- Našli **137 nových polymérov** s požadovanými vlastnosťami (**z pôvodných 4**).
- Použili **100 generácií** a kumulatívne otestovali **12675 polymérov**.
- Tým, že vysoká tepelná stabilita a schopnosť zniesť vysoké elektrické pole sú intervaly hodnôt, autori uviedli, že rátajú s nejakou mierou chyby, keďže fitness funkcia bol ML prediktor.



Zdroj

- Chiho Kim, Rohit Batra, Lihua Chen, Huan Tran, Rampi Ramprasad,
Polymer design using genetic algorithm and machine learning,
Computational Materials Science,
Volume 186,
2021,
110067,
ISSN 0927-0256,
<https://doi.org/10.1016/j.commatsci.2020.110067>.
(<https://www.sciencedirect.com/science/article/pii/S0927025620305589>)

Ďakujeme za pozornosť

Priestor pre Vaše otázky a názory