

# A Hyper Learning Binary Dragonfly Algorithm for Feature Selection: A COVID-19 Case Study [1]

Rešerše článku - Martin Juříček

Faculty of Mechanical Engineering, Brno University of Technology  
Institute of Automation and Computer Science  
Technická 2896/2, Brno 616 69, Czech Republic  
200543@vutbr.cz

**Abstrakt:** *Tato rešerše článku pojednává a stručně shrnuje algoritmus Hyper Learning Binary Dragonfly Algorithm s případovou studií Covid-19.*

**Klíčová slova:** *Biologicky inspirovaný algoritmus, optimalizační algoritmus, Hyper Learning Binary Dragonfly Algorithm*

## 1 Úvod

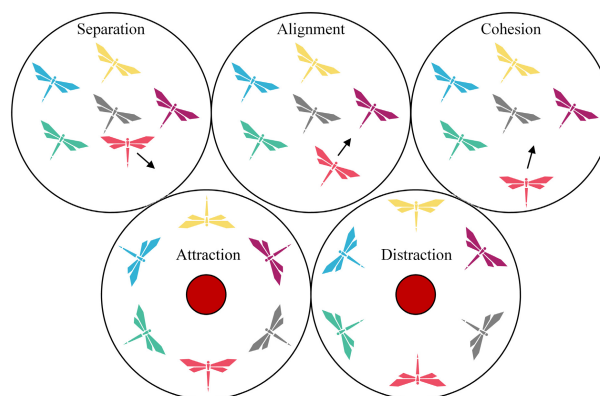
V dnešní době se čím dál tím častěji skloňuje pojem umělá inteligence. Tento pojem sebou nese oblasti jako je například machine learning(strojové učení) nebo data mining(vytěžování dat). Výběr prvků(feature selection) je typickým a klíčovým úkolem při využívání metod data mining v části procesu předzpracování. Hlavním úkolem je nalezení optimální podmnožiny prvků z dané množiny všech prvků. K řešení selekce prvků se ukázaly jako velmi silným a schopným nástrojem metaheuristiky.

V posledním desetiletí jsou metaheuristiky považovány za neefektivnější a nejspolehlivější optimalizační techniky zvláště při řešení náročných, vysokodimenzionálních problémů, které může i představovat zpracování dat, a s ním extrakce prvků. Metaheuristiky zahrnují genetic algorithms(genetické algoritmy), water cycle algorithm(algoritmus koloběhu vody) nebo lightning search algorithm (algoritmus pro vyhledávání blesků). Do této sekce lze zařadit i dragonfly algorithm(algoritmus vážky).

## 2 Hyper Learning Binary Dragonfly Algorithm

Hyper Learning Binary Dragonfly Algorithm(HLBDA) je rozšířenou verzí původního Dragonfly Algoritmu(DA). Algoritmus byl publikován v roce 2016, přičemž byl navržen k řešení spojitých optimalizačních úloh a je inspirován vzory rojení vážek v přírodě. Současně existuje varianta pro diskrétní úlohy Binary Dragonfly algorithm(BDA). Ačkoliv tato verze pro diskrétní systémy prokázala svou výkonnost v několika srovnávacích testech, nicméně slabá stránka se může objevit u problému lokálního optima.

Explorační a exploatační mechanismy DA jsou modelovány podle interakce vážek při vyhýbání se a hledání zdroje potravy. V DA existuje pět hlavních způsobů chování, a to oddělování, vyrovnávání, soudržnost, přitahování, a rozptýlení, která se používají v procesu aktualizace polohy(obr. 1).



Obrázek 1: Oddělování, vyrovnávání, soudržnost, přitahování a rozptýlení

HLBDA je navržen s cílem aplikování na řešení úkolů selekce prvků. Tento algoritmus využívá strategii hyperlearning(hyperučení), která zahrnuje koncept osobního nejlepšího a osobního nejhoršího řešení v dané pozici(obr. 1). V konvenční BDA se vázky soustředí na globálně nejlepší řešení(zdroj potravy) a globálně nejhorší řešení(nepřítel) pro přilákání a odvedení pozornosti.

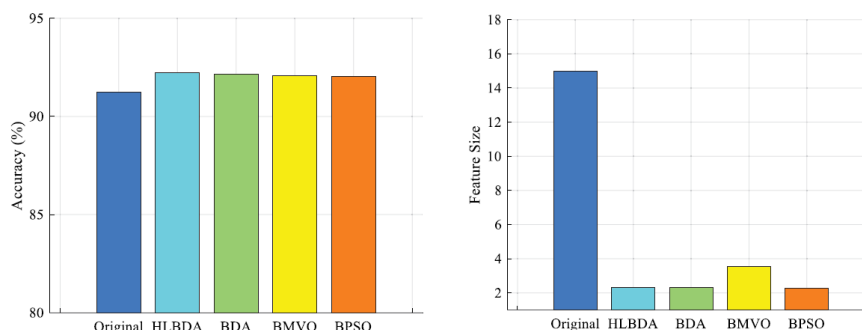
V tomto případě však platí, že přidáním osobního nejlepšího a osobního nejhoršího řešení vázky do definovaných aktivit, předpokládá, že potenciál hledání potravy se zvýší a chování při útěku před nepřítelem lze zlepšit. Strategie hyperlearning navíc umožňuje vázkám se učit jak z osobních nejlepších, tak z globálně nejlepších řešení během fáze vyhledávání. Místo aktualizace polohy na základě chování v roji, se vázka snaží zlepšit od svých osobních nejlepších a globálních nejlepších zkušeností.

### 3 Srovnání s jinými metodami

V rámci této práce byl porovnán algoritmus HLBDA s osmi dalšími optimalizačními algoritmy, které lze použít pro řešení extrakce prvků. Z komparačních testů vyplývá, že HLBDA nacházel optimální nejlepší fitness(shodu) na sedmnácti datasets(souborech dat). Dobrá vyhledávací schopnost HLBDA při řešení selekce prvků těží především z navržené strategie hyperlearning. Při řešení tohoto úkolu HLBDA může často poskytovat vysoce konzistentní výsledky ku porovnání s BDA. Na základě získaných výsledků lze konstatovat také, že klasifikační schopnost HLBDA byla podstatně lepší než u ostatních algoritmů. Ve výsledku navržená metoda HLBDA nejenže nabízí velmi vysokou přesnost klasifikace, ale také dobrou redukci dimenzionality při řešení selekce prvků.

### 4 Případová studie Covid-19

Svého implementačního nasazení se algoritmus HLBDA dočkal pro řešení případové studie předpovědi zdravotního stavu pacienta na datasetu sesbíraného v rámci pandemie Covid-19. Tento dataset lze získat na serveru github: <https://github.com/Atharva-Peshkar/Covid-19-Patient-Health-Analytics>. Dataset se skládá z patnácti prvků, přičemž záměrem této studie je předpovídat stavy úmrtí a uzdravení v závislosti na daných faktorech. Na obrázku 2 lze vidět přesnost a velikost prvků. HLBDA dosáhl nejvyšší klasifikační přesnosti 92,21 % s využitím pouze tří prvků z databáze.



Obrázek 2: Přesnost a velikost prvků navrhovaného HLBDA na souboru dat COVID-19

### 5 Závěr

Své aplikační opodstatnění získávají biologicky inspirované algoritmy především jako optimalizační algoritmy. Díky využití těchto algoritmů lze zlepšovat oblasti umělé inteligence jako je například data mining. Příkladem může být aplikace HLBDA pro extrakci prvků při řešení případové studie predikce stavů pacienta z databáze získané během pandemie Covid-19.

### Odkazy

- [1] TOO, Jingwei; MIRJALILI, Seyedali. A Hyper Learning Binary Dragonfly Algorithm for Feature Selection: A COVID-19 Case Study. *Knowledge-Based Systems*. 2021, **212**, 106553. ISSN 0950-7051. Dostupné z DOI: <https://doi.org/10.1016/j.knosys.2020.106553>.