FlashQuation: Vergelijkingen vinden aan de hand van voorbeelden

Jeroen Craps & Tom De Groote

Introductie

Probleemgebied:

De slechte gebruikerservaring bij het zoeken naar een passende vergelijking voor een probleem.

Onderzoeksvraag:

Is het mogelijk om in beperkte tijd een vergelijking te vinden in de door gebruiker gedefinieerde set van getallen?

Hypothese:

Het is mogelijk om een vergelijking te vinden in beperkte tijd over een gegeven set getallen.

Methoden

STAPPEN:

1. Eenmalig de bewerkingsboom bereken aan de hand van CFG.
2. Verwijderen redundante knopen.
   → Optimale bewerkingsboom.
3. Voorbeeld wordt gegeven.
4. Efficiënt zoeken van vergelijkingen die voldoen aan het gegeven voorbeeld.

Hypothese 1:

Er is een significant verschil tussen het aantal knooppunten van een optimale boom en van de volledige boom.

Hypothese 2:

Er kunnen meer unieke oplossingen gevonden worden in beperkte tijd met behulp van de optimale boom.

Voor de zoekruimte te verbreden worden constanten toegevoegd aan de CFG.

Hypothese 3:

Door middel van deze constanten neemt de algemene oplossingsgraad toe.

Om de oplossingsgraad te verhogen worden geen overbodige vergelijkingen uitgerekend. (vb. T0*T1/T1 ~ T0)

Hypothese 4:

Het efficient evalueren verhoogt de oplossingsgraad ten opzichte van de brute-force manier.

Contextvrije grammatica (CFG)

Productieregels van de vorm:

E → T **(R1)**
E → E O T **(R2)**
T → 1..9 I

met E een niet-terminaal symbool, T een terminaal symbool en O een operand.

Voorbeeld:

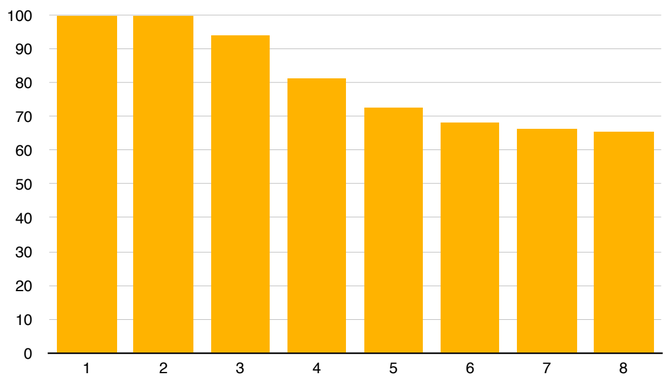
E
→ **E O T**
→ **T O T**

T wordt bij het evalueren ingevuld door gegeven waarde of gewichten.

Resultaten

Hypothese 1:

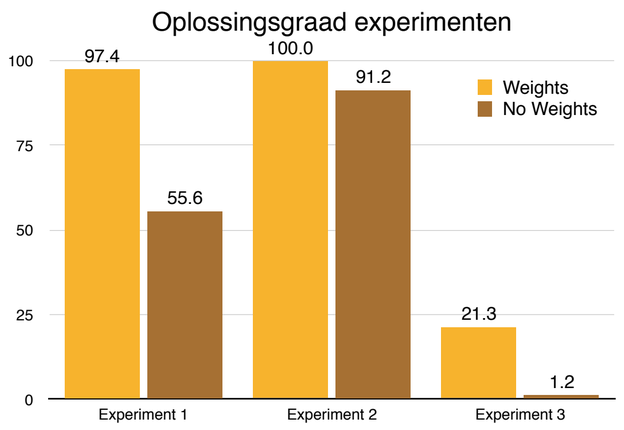
Naarmate de diepte van de boom toeneemt worden er meer redundante knopen verwijderd.



Procentuele weergave van het aantal knooppunten van de optimale boom in vergelijking met de volledige boom.

Hypothese 2:

Verwacht wordt dat de tijdswinst dezelfde curve zal volgen.



E1: Random met gekende vergelijking met 3 geg. vb'en
E2: Random zonder gekende vergelijking met 1 geg. vb'en.
E3: Random zonder gekende vergelijking met 2 geg. vb'en.

Hypothese 3:

De oplossingsgraad met gewichten is duidelijk hoger in alle experimenten. Verwacht wordt dat de combinatie van E1 en E2 een goede representatie is van de door de gebruiker gezochte vergelijking.

Conclusie

Door het berekenen van de optimale contextvrije boom (met een beperkt aantal operaties) en het gebruiken van gewichten kunnen reeds in ongeveer 70% van de gevallen vergelijkingen gevonden worden als er geen tijdsbeperking is. Het is van uiterst belang dat de zoekruimte efficiënt zal doorlopen worden om deze geschatte oplossingsgraad te benaderen.

Toekomstig onderzoek:

- De correlatie tussen de oplossingsgraad en welke constanten (of operanden) gebruikt worden.
- Implementeren van een efficiënt evaluatiealgoritme. **(Hypothese 4)**

Referenties:

- Spreadsheet Table Transformations from Examples, CACM, W.R. Harris, S. Gulwani
- Example-Based Learning in Computer-Aided STEM Education, CACM 2014, Sumit Gulwani
- Equation Discovery, Enclyclopedia of Machine Learning

Begeleiders: Prof. L. De Raedt
Postdoc. A. Kimmig

Contact: jeroen.craps@student.kuleuven.be
tom.degroote@student.kuleuven.be

KU LEUVEN