

mgr Tomasz Peszek

Promotor: dr hab. Janusz Jurek (AGH) **Dyscyplina:** Informatyka

Problematyka syntaktycznego rozpoznawania wzorców zniekształconych/rozmytych ma duże znaczenie. Dzięki swoim właściwościom (możliwość analizy złożonych obiektów strukturalnych, możliwość opisu nieskończonych klas obiektów za pomocą skończonej liczby reguł) metody syntaktycznego rozpoznawania wzorców znalazły liczne zastosowania praktyczne m.in. w rozpoznawaniu pisma oraz mowy, w biometrii, w analizie i przetwarzaniu obrazów wizyjnych, w analizie kształtów, w analizie grafów, w analizie danych, w prognozowaniu, w klasyfikacji tekstów, w systemach wspomagających diagnostykę medyczną, w systemach naprowadzania i nawigacji, w systemach kontroli procesów produkcyjnych oraz w wielu innych. Z aplikacjami syntaktycznych metod rozpoznawania wzorców są związane pewne trudności m.in. duża wrażliwość na jakość danych wejściowych, w tym ich zniekształcenia lub rozmycie – nawet niewielkie zniekształcenie danych wejściowych może doprowadzić do niepowodzenia procesu analizy całego wzorca. W takiej sytuacji celowe wydaje się zdefiniowanie nowego syntaktycznego modelu rozpoznawania wzorców uwzględniającego aspekt rozmycia/zniekształcenia analizowanych wzorców.

We „Wstępie” do rozprawy (rozdział 1), na podstawie przeprowadzonego badania literatury przedmiotu syntaktycznego rozpoznawania wzorców, został opisany kontekst badań, w tym główne kierunki badawcze oraz najważniejsze otwarte problemy. W ramach przedstawionego kontekstu zdefiniowano i osadzono problem naukowy, którego rozwiązanie było przedmiotem badań własnych m.in. uzasadniono konieczność stworzenia nowego syntaktycznego modelu rozpoznawania wzorców rozmytych/zniekształconych oraz sformułowano tezę dysertacji w następującej postaci (strona 16): *„Możliwe jest opracowanie nowego modelu syntaktycznego rozpoznawania wzorców uwzględniającego aspekt rozmycia/zniekształcenia wzorców, który posiada istotne zalety w porównaniu z innymi, dotychczas stosowanymi modelami syntaktycznego rozpoznawania wzorców.”*

Teza pracy została wykazana w następujący sposób:

- 1) *Zdefiniowano syntaktyczny model rozpoznawania wzorców zniekształconych w oparciu o nową klasę analizatorów składniowych (automat FDPLLA(k)).*

- 2) Wykonano implementację aplikacji realizującej model automatu $FDPLLA(k)$.
- 3) Implementację zweryfikowano na przykładach prostych języków formalnych oraz poprzez jej wykorzystanie w systemie generującym korektę prognozy zapotrzebowania na energię (na danych udostępnionych przez Tauron PE S.A.).
- 4) Dokonano porównania zdefiniowanego syntaktycznego modelu rozpoznawania wzorców rozmytych/zniekształconych wykorzystującego automat klasy $FDPLLA(k)$ z syntaktycznymi modelami stosowanymi dotychczas. Wykazano istotne zalety nowego modelu.

Teza rozprawy została wykazana w kolejnych rozdziałach pracy.

Rozdział drugi został poświęcony wprowadzeniu pojęć niezbędnych dla zdefiniowania nowego, syntaktycznego modelu rozpoznawania wzorców rozmytych/zniekształconych. Spośród najważniejszych zdefiniowanych pojęć trzeba wymienić pojęcia: prymitywu rozmytego, składowej prymitywu rozmytego, elementu prymitywu rozmytego, wymiaru prymitywu rozmytego, słowa rozmytego, rozwiązania słowa rozmytego, długości słowa rozmytego. Równocześnie, w rozszerzeniu klasycznego pojęcia wyniku analizy syntaktycznej wzorca, określono definicję wyniku analizy syntaktycznej słowa rozmytego.

W rozdziale trzecim zdefiniowano formalnie nowy, syntaktyczny model rozpoznawania wzorców rozmytych/zniekształconych, w szczególności podano dokładny opis konstrukcji i zasady działania nowej klasy automatów tj. $FDPLLA(k)$ (*Fuzzy Dynamically Programmed $LL(k)$ Automaton*) umożliwiających efektywną analizę słów rozmytych. Szczegółowo opisano architekturę automatu $FDPLLA(k)$ wraz z jej najważniejszymi elementami. Model automatu $FDPLLA(k)$ został oparty na gramatykach klasy $DPLL(k)$ ze względu na ich dużą siłę dyskryminacyjną oraz zdefiniowane, wielomianowe algorytmy analizy syntaktycznej, inferencji gramatyki oraz konstrukcji parsera. Na podstawie zdefiniowanego modelu formalnego oszacowano złożoność działania automatu $FDPLLA(k)$ – zarówno złożoność czasową jak i pamięciową.

W rozdziale trzecim sformułowano także ogólne wytyczne dotyczące możliwości zrównoleglenia obliczeń w trakcie analizy syntaktycznej prowadzonej z użyciem automatu $FDPLLA(k)$. Kończącą sekcję rozdziału poświęcono porównaniu nowej metody

analizy syntaktycznej z głównymi, syntaktycznymi podejściami stosowanymi dotychczas w rozpoznawaniu wzorców rozmytych/zniekształconych (gramatyki stochastyczne oraz gramatyki z korekcją błędów). Podsumowano zalety nowego modelu tj. możliwość analizy syntaktycznej słów rozmytych, wielomianową złożoność obliczeniową algorytmu analizy przy zachowaniu dużej siły opisowej (dzięki konstrukcji automatu $FDPLLA(k)$ oraz wykorzystaniu gramatyk klasy $DPLL(k)$), brak konieczności klasyfikacji typów zniekształceń *a priori* oraz istnienie algorytmu inferencji gramatyki (ułatwia praktyczne zastosowanie modelu).

W rozdziale czwartym opisano implementację aplikacji realizującej model automatu $FDPLLA(k)$. Na bazie zaimplementowanej aplikacji przeprowadzono testy z wykorzystaniem prostych języków kształtów. Testy potwierdziły odporność zdefiniowanego modelu analizy syntaktycznej na zniekształcenia danych wejściowych oraz dobre właściwości obliczeniowe.

Należy podkreślić, że model $FDPLLA(k)$ został również wstępnie zweryfikowany w zastosowaniach praktycznych. W rozdziale 5 opisano wykorzystanie modelu automatu $FDPLLA(k)$ w systemie generującym krótkoterminowe prognozy zapotrzebowania na energię. Prototyp systemu powstał we współpracy z firmą *Tauron Polska Energia S.A.* oddział w *Gliwicach*. System został zweryfikowany na danych realnych dotyczących wybranego obszaru dystrybucyjnego za okres dwóch lat. Dane dotyczące pierwszego roku utworzyły zbiór uczący dla aplikacji. Skuteczność systemu została zweryfikowana na danych dotyczących drugiego roku. Zaimplementowany system osiągnął na danych walidujących wyniki lepsze od systemów prognozowania używanych aktualnie w przedsiębiorstwie (metody oparte na sieciach neuronowych, autoregresji oraz wiedzy eksperckiej). Wynik ten, jakkolwiek bardzo obiecujący, wymaga dalszych weryfikacji na szerszym zbiorze danych.

W rozdziale 6 zawarto krótkie podsumowanie wyników badań oraz referat sposobu wykazania postawionej tezy rozprawy.

W dodatkach A i B zebrano najważniejsze pojęcia z dziedziny syntaktycznego rozpoznawania wzorców oraz języków formalnych. W dodatku C podano ogólne wytyczne dotyczące konstrukcji automatu klasy $FDPLLA(k)$ dla $k > 1$.

Do głównych wyników osiągniętych w toku badań własnych należy zaliczyć:

1. Rozszerzenie modelu syntaktycznego rozpoznawania wzorców w oparciu o definicję prymitywów rozmytych i słów rozmytych.
2. Zdefiniowanie nowej klasy automatów $FDPLL(k)$, umożliwiających efektywną obliczeniowo analizę słów rozmytych.
3. Zastosowanie opracowanego modelu do systemu prognozującego zapotrzebowanie na energię elektryczną.
4. Porównanie nowego syntaktycznego modelu rozpoznawania wzorców zniekształconych/rozmytych ze stosowanymi dotychczas metodami syntaktycznego rozpoznawania wzorców zniekształconych/rozmytych oraz wykazanie zalet nowego modelu.

Obok wymienionych powyżej podstawowych osiągnięć, warto wskazać również istotne wyniki szczegółowe, takie jak: opracowanie metody generowania prymitywów rozmytych w oparciu o probabilistyczne sieci neuronowe oraz wstępną weryfikację możliwości implementacji modelu w środowisku równoległym.