



Wydział Matematyki i Nauk Informatycznych
kierunek informatyka, specjalność informatyka stosowana

Andrzej Wąsowski

Zdalna generacja sieci bayesowskich z baz danych

Praca dyplomowa pod kierunkiem
dr hab. inż. Mieczysława A. Kłopotka

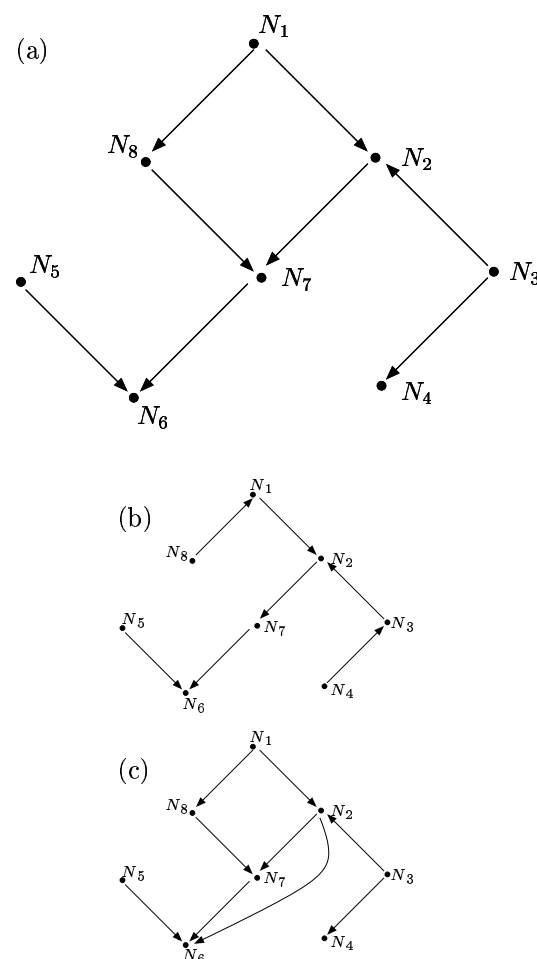
Automatyczne i półautomatyczne wnioskowanie z wykorzystaniem baz wiedzy jest dziś jednym z najistotniejszych zastosowań sztucznej inteligencji. Jego znaczenie rośnie wraz z rozwojem kręgu użytkowników systemów eksperckich, czyli różnego rodzaju narzędzi wspomagających podejmowanie decyzji. Analiza danych mikro i makroekonomicznych, testowanie wiarygodności kredytowej, diagnostyka chorób, klasyfikacja dokumentów, czy para-inteligentny system pomocy z usuwaniem problemów przy drukowaniu w systemie Microsoft Windows to tylko niektóre z licznych zastosowań takich systemów, niejednokrotnie opartych o zasady analizy bayesowskiej.

Sieci przyczynowo-skutkowe, zwane też sieciami Bayesa, należą do często stosowanych sposobów reprezentacji w bazach wiedzy systemów eksperckich. Sieć taka reprezentuje układ zależności między różnymi parametrami w postaci grafu skierowanego. Generację sieci nazywa się często uczeniem lub odtwarzaniem.

W pracy rozważono zagadnienie zdalnej generacji sieci Bayesa. Przedstawiono pojęcia podstawowe z zakresu sieci przyczynowo-skutkowych i zasady funkcjonowania poszczególnych algorytmów uczących. Następnie zrealizowano demonstracyjny system uczący o nazwie *RENEG*, który pozwolił wykonać testy porównawcze dla dziewięciu różnych algorytmów generacji.

Zastosowano twórcze podejście do implementacji istniejących algorytmów, niejednokrotnie stosując własne usprawnienia w postaci równoważnych przekształceń oryginalnych sformułowań funkcji i metod postępowania, inwersję reprezentacji, wymianę używanych przez autorów algorytmów podstawowych (niższego poziomu) na alternatywne skuteczniejsze algorytmy, ingerowanie w kolejność podawania podzbiorów do testów niezależności, oraz stosując inżynierskie usprawnienia w rodzaju zamiany licznych mnożeń zmiennoprzecinkowych na sumy logarytmów, czy tablicując funkcje pomocnicze. W niektórych al-

gorytmach modyfikowano nieznacznie warunki stopu, inne z kolei wymagały doszczegółowienia elementów mniej kluczowych, ale niezbędnych do funkcjonowania i nieokreślonych przez autorów. Zapropnowano też własny mechanizm zliczania wartościowań zmiennych i grup zmiennych w próbie oparty o zagnieżdżane drzewa pozycyjne.



Przykładowa sieć przyczynowo skutkowa: a) pozostać oryginalna, b) i c) próby odtworzenia sieci z wykorzystaniem algorytmów uczących systemu *RENEG*.

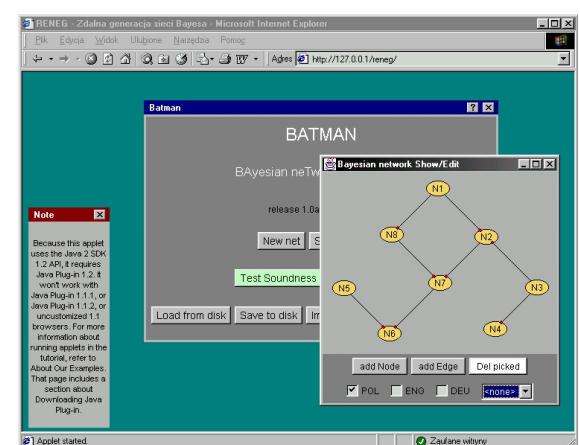
Zrealizowany w ramach pracy system zdalnej generacji *RENEG* intensywnie korzysta z technologii internetowych. Nikt nie ma już dziś wątpliwości, że Internet stał się motorem rozwoju nie tylko informatyki, ale i całej gospodarki. Rozwój sieci globalnej przerósł wszelkie oczekiwania, a łatwa dostępność do Internetu w wielu regionach sprawiła, że stał się on najtańszym i najskuteczniejszym medium dys-

trybucji produktów o charakterze elektronicznym (w tym oprogramowania).

Zaimplementowany system został zbudowany w oparciu o architekturę klient/serwer, gdzie serwer WWW pełni rolę serwera aplikacji. Strona klienta oparta jest na przeglądarce WWW a części systemu komunikują się przez Internet za pośrednictwem technologii CGI. System cechuje też prosty i łatwy w użyciu graficzny interfejs użytkownika oraz bogata dokumentacja: pomoc kontekstowa, instrukcja użytkownika i instrukcja eksploatacyjna.

Dzięki zastosowaniu technologii internetowych system *RENEG* zyskuje wiele na elastyczności. Praktyczny brak wymagań co do systemu operacyjnego i platformy sprzętowej klienta rozszerza bardzo zakres użytkowników. Zastosowane technologie internetowe (http, html, Java i JavaScript) cechuje daleko posunięta przenośność między platformami.

Nie ograniczając się do zapewnienia przenośności strony klienta, autor posunął się krok dalej. Zastosowano techniki programowania przenośnego w językach C/C++, co zaowocowało utworzeniem systemu przenośnego także ze względu na platformę serwera. Dzięki temu stało się możliwe porównywanie własności obliczeniowych różnych systemów operacyjnych.



Ta sama sieć przedstawiona w aplikacji wizualizacyjnej BATMAN (dołączonej do projektu).