

SPRAWOZDANIE Z DZIAŁALNOŚCI NAUKOWO-BADAWCZEJ KATEDRY METOD KOMPUTEROWYCH UMK W ROKU 2002

I. TEMATY BADAWCZE

1. Liczba realizowanych tematów.

- 14, w tym 10 w ramach działalności statutowej i 4 w ramach badań własnych.

2. Liczba tematów zakończonych: 0 - w planowanym terminie 0, - z opóźnieniem 0.

- Prowadzone badania mają w większości charakter długoterminowy.

3. Z liczby tematów zakończonych przekazano:

- do wykorzystania, - do wdrożenia.

4. Z liczby tematów przekazanych do wdrożenia, w 2002 r. wdrożono:

- ogółem 1, - w tym w resorcie edukacji narodowej 0.

- Wdrożenie dotyczy metod inteligencji obliczeniowej zawartych w pakiecie programów GhostMiner do data mining, opracowanym przy współpracy z firmą FQS Poland (oddział Fujitsu). Nie jest to temat zakończony, gdyż powstają kolejne wersje systemu w oparciu o nowsze badania.

5. Liczba tematów, których realizację zaniechano (w roku sprawozdawczym).

- 1 temat, w 2002 roku nie zajmowaliśmy się aktywnie neuroinformatyką, ale mamy nadzieję powrócić do tego tematu.

Liczba realizowanych tematów ogółem		14
Działalność statutowa		10
Badania własne		4
Granty UMK		0
Liczba tematów zakończonych		
w tym	W planowanym terminie	0
	z opóźnieniem	0
z liczby tematów zakończonych przekazano	Do wykorzystania	0
	Do wdrożenia	0
z liczby tematów do wdrożenia, w 2002 wdrożono	Ogółem	1
	W tym w resorcie edukacji narodowej	0
Liczba tematów, których realizację zaniechano		1

II. CZĘŚĆ OPISOWA

II.1. Przyczyny zaniechania tematów prowadzonych i rozpoczętych w 2002 roku.

W 2002 roku nie zaniechaliśmy realizacji żadnego tematu.

II.2. Krótka charakterystyka i podsumowanie wyników uzyskanych w 2002 roku.

Problematyka badawcza Katedry Informatyki Stosowanej skoncentrowana jest wokół czterech grup tematycznych.

A) Inteligencja obliczeniowa (computational intelligence)

Zagadnienia inteligencji obliczeniowej dotyczą teorii i zastosowań modeli sieci neuronowych, logiki rozmytej, drzew decyzyjnych, algorytmów opartych na precedensach (case-based), oraz fundamentalnych mechanizmów działania struktur adaptujących się, w tym również zagadnień związanych z działaniem mózgu i modelowaniem umysłu. Jest to część interdyscyplinarnego programu badawczego kierowanego przez prof. W. Duchą, finansowanego w części dotyczącej rozwoju i zastosowań algorytmów inteligencji obliczeniowej przez zakończony w 2002 roku grant KBN „Teoria i zastosowania metod inteligencji obliczeniowej”, którego głównym celem było stworzenie narzędzi do data mining. Współpracujemy też z oddziałem firmy Fujitsu w Polsce, FQS Poland, nad wykorzystaniem opracowanych przez nas algorytmów w systemach data mining.

Główne zadania badawcze należące do tej grupy tematycznej to:

II.2.1. Odkrywanie wiedzy w bazach danych za pomocą metod inteligencji obliczeniowej.

Jest to część bardzo modnej obecnie dziedziny, określanej jako „data mining”, czyli dogłębna analiza danych lub odkrywanie wiedzy w bazach danych. W roku 2002:

- Ulepszono metody ekstrakcji reguł oparte na kryterium separowalności (SSV), prowadzące do sprawnego algorytmu typu drzewa decyzyjnego. Testowano zastosowanie tego kryterium w połączeniu z funkcjami odległości, znajdując bardzo proste reguły prototypowe. Zastosowano również oryginalne podejście do tworzenia „lasu drzew decyzyjnych”, o podobnej dokładności, ale różniących się między sobą wykorzystywanymi cechami jak i dokładnością (czułością i wrażliwością) dla specyficznych klas. Jest to część pracy doktorskiej Krzysztofa Grąbczewskiego, która powinna zostać obroniona w połowie 2003 roku.
- Reguły logiczne wymagają wstępnej dyskretyzacji danych. Zastosowaliśmy w tym celu opracowane przez nas kryterium separowalności (SSV, Separability Split Value), pozwalające budować drzewa decyzyjne lub tworzyć optymalne podziały ciągłych zmiennych i wybierać podzbiory zmiennych dyskretnych. Kryterium to można zastosować zarówno dla szukania najlepszego sposobu dyskretyzacji jak i tworzenia zestawów reguł. Opracowano specjalną wersję programu do celów dyskretyzacji i pokazano, że informacja wzajemna (mutual information), obliczona pomiędzy klasami i cechami, jest dla realnych zbiorów danych ponad dwukrotnie większa niż w przypadku prostych metod dyskretyzacji, opartych na stałej liczbie wektorów lub wielkości przedziałów.
- Pracowaliśmy nad metodami szukania charakterystycznych prototypów wykorzystując probabilistycznie określone miary podobieństwa. Pozwala to na wprowadzenie nowego rodzaju reguł opartych na podobieństwie do prototypów. Opracowano metodę generacji funkcji przynależności logiki rozmytej odpowiadających różnym metrykom. Prowadzi to do tworzenia funkcji przynależności na podstawie danych, zamiast powszechnie stosowanej metody przyjmującej *a priori* proste funkcje przynależności, których parametry są następnie optymalizowane. Przyjęcie miary odległości w formie addytywnych przyczynków zależnych od pojedynczych cech (przykładem jest miara odległości

Manhattan) pozwala na bezpośrednią interpretację takich reguł w sensie logiki rozmytej. Reguły klasyczne otrzymać można dla używając metryki Chebyscheva. Reguły tego typu generować można będzie za pomocą wielu metod inteligencji obliczeniowej: sieci neuronowych, drzew decyzyjnych i metod opartych na podobieństwie. Kolejnym krokiem będzie zbadanie możliwości optymalizacji prototypów przez optymalizację ich własności.

- Testowaliśmy i rozwijaliśmy metody generacji reguł oparte na zastosowanie algorytmów szukania wiązką lub szukania typu „pierwszy najlepszy” dyskretnych wektorów parametrów opisujących sieć neuronową. Udało się w ten sposób otrzymać interesujące reguły. Ponieważ do oceny perturbacji parametrów wykorzystuje się tylko informację o liczbie błędów jest to podejście ogólne, szczególnie przydatne w poszukiwaniu ostrych reguł logicznych, gdy obszar niezerowych gradientów kurczy się do zera i sieci z wsteczną propagacją błędów nie uczą się.
- Kontynuowano współpracę z firmą FQS Poland (Fujitsu Group) zainteresowaną stworzeniem nowego systemu do dogłębnej analizy danych (data mining), opartego na naszych algorytmach. Wersja 1.0 systemu została ukończona i jest obecnie sprzedawana przez FQS. Wersję demonstracyjną pobrało kilkaset osób. Jest to udane wdrożenie, które pozwoliło również członkom naszej grupy na łatwiejsze porównywanie i integrację nowych algorytmów, ułatwiając korzystanie z poszczególnych, rozwijanych dotychczas niezależnie programów.
- W. Duch napisał obszerną pracę na ten temat na zamówienie *Proceedings of IEEE*, najbardziej prestiżowego pisma IEEE (współautorami są R. Setiono, Singapore i J. Żurada, Louisville).
- W. Duch wygłosił 3-godzinny tutorial „Computational Intelligence for Data Mining”, na najważniejszej konferencji w 2003 roku, World Congress on Computational Intelligence, Honolulu, USA, 17-24.05.2002. Był też wykładowcą na szkole europejskiej “Neural networks in supervised classification and regression” w Porto, Portugalia, 8-12.07.2002, prowadząc tam dwa moduły (po 2 godziny każdy): „Data mining I” oraz „Data Mining - the fuzzy way”.

Temat ten rozwijany jest przy współpracy z prof. M. Ishikawą z Kyushu Institute of Technology (podpisaliśmy formalną umowę o współpracy z Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Japonia), prof. Y. Hayashi z Meiji University, Tokyo, i prof. R. Setiono z National University of Singapore.

Publikacje związane z tym tematem:

1. Grąbczewski K, Duch W, Forests of decision trees. International Conference on Neural Networks and Soft Computing (ICNNSC), Advances in Soft Computing, Physica Verlag (Springer) 2002, pp. 602-607
2. Grąbczewski K, Duch W, Heterogenous forests of decision trees. Springer Lecture Notes in Computer Science Vol. 2415, pp. 504-509.
3. Duch W, Grąbczewski K, Heterogeneous adaptive systems. IEEE World Congress on Computational Intelligence, Honolulu, May 2002, pp. 524-529.
4. Duch W, Setiono R, Zurada J, Computational intelligence methods for understanding of data. Proc. of the IEEE (wysłane 12/2002.)
5. Grudzinski K, „Similarity-based methods in application to analysis of scientific and medical data”, Praca doktorska, KIS UMK, Toruń 2002.
6. Duch W, GhostMiner Developer tutorial; wyd. FQS Poland, CD-ROM, łącznie z programem (9 str. druku).
7. Grąbczewski K, Jankowski N, GhostMiner Developer and Analyzer Help files, wyd. FQS Poland, CD-ROM, łącznie z programem (ok 150 str. druku)
8. Duch W, Adamczak R, Grąbczewski K, Grudziński K, Jankowski N, Naud N, Computational Intelligence for Data Mining. In: World Congress on Computational Intelligence, Honolulu, USA 17-24.05.2002 (tutorial, 50 str).

Szukanie wiedzy w bazach danych (data mining) nabiera coraz większego znaczenia i na wielu konferencjach organizuje się sesje specjalne poświęcone temu zagadnieniu. Temat ten będzie przez naszą grupę nadal intensywnie rozwijany.

II.2.2. Rozwój teorii i zastosowań systemów neurorozmytych.

Rozwijany przez nas od kilku lat model neurorozmyty Feature Space Mapping (FSM) estymuje prawdopodobieństwa rozkładu danych za pomocą sieci elementów, które realizują funkcje separowalne. Pozwala to na interpretację tych funkcji w jako funkcji przynależności w sensie logiki rozmytej. Jest to model bardzo bogaty, można go rozpatrywać z wielu punktu widzenia, jako sieć neuronową, system logiki rozmytej korzystający z zależnych od kontekstu funkcji przynależności, system oparty na podobieństwie (przy użyciu funkcji zlokalizowanych) integrujący dwa podstawowe modele kategoryzacji w psychologii (teorię prototypów i teorię egzemplarzy), a nawet jako system realizujący heurystyki w procesach poszukiwania rozwiązań.

W 2002 roku zajmowaliśmy się głównie dostosowaniem oprogramowania FSM do pakietu GhostMiner, służącego do dogłębnej analizy danych. Wymagało to opracowania odpowiedniej struktury klas w języku C++ i interfejsów do dostępnych funkcji pakietu GhostMiner, np. dotyczących tworzenia komitetów sieci czy sposobów testowania. Zastosowano szereg drobnych usprawnień. FSM stosowany jest w ramach tego pakietu zarówno do klasyfikacji jak i do ekstrakcji reguł logicznych.

II.2.3. Ontogeniczne sieci neuronowe i systemy heterogeniczne.

Sieci ontogeniczne dopasowują swoją złożoność do danych pozwalając na znalezienie nieobciążonych modeli klasyfikatorów o minimalnej wariancji. Systemy heterogeniczne używają różnych elementów w celu odkrycia najprostszyc struktur w danych. Naszym celem jest połączenie możliwości tych dwóch rodzajów sieci w celu stworzenia modeli działających automatycznie, bez konieczności doboru architektury sieci, tworzących modele o minimalnej złożoności.

W ramach prac nad tymi zagadnieniami w 2002 roku dr Jankowski rozwinął oprogramowanie dla sieci IncNet realizujące algorytm Optimal Transfer Functions. Zagadnienie to okazało się jednak technicznie trudne i wymaga dłuższej pracy. Chociaż mamy już pomysł na efektywny algorytm dla uczenia tego typu sieci nie został on jeszcze przetestowany.

Dr Jankowski przygotował monografię na temat sieci ontogenicznych, która zostanie wydana w 2003 roku, obecnie trwają dyskusje w sprawie finansowania jej wydania.

Ideę heterogenicznych sieci neuronowych wykorzystujących funkcje transferu różnego typu rozwinęliśmy szerzej opisując heterogeniczne drzewa decyzji i systemy oparte na podobieństwie. Przeprowadzono wstępne testy dla heterogenicznych drzew decyzji osiągając bardzo obiecujące wyniki. Jest tu nadal dużo wyzwań i otwartych problemów, niemniej dotychczasowe wyniki są bardzo zachęcające.

Publikacje związane z tym tematem

1. Grąbczewski K, Duch W, Heterogenous forests of decision trees. Springer Lecture Notes in Computer Science Vol. 2415, pp. 504-509.
2. Duch W, Grąbczewski K, Heterogeneous adaptive systems. IEEE World Congress on Computational Intelligence, Honolulu, May 2002, pp. 524-529.

II.2.4. Zastosowanie metod inteligencji obliczeniowej do analizy danych.

W roku 2002 przeanalizowaliśmy kilka medycznych baz danych.

- Korzystając z publicznie dostępnych baz danych porównywaliśmy wiele systemów klasyfikujących dane jak i systemów dokonujących ekstrakcji danych. Wyniki tych badań są ciągle uzupełniane i udostępniane naszych stronach internetowych.
<http://www.phys.uni.torun.pl/KIS/projects/datasets.htm>
- Otrzymaliśmy nowe rezultaty dla danych medycznych oraz danych dotyczących datowania obiektów archeologicznych (z Austrii) na podstawie analizy spektrograficznej próbek znalezionych szkieł. Odkryto interesujący problem związany z obecnością kilku próbek z tego samego obiektu w bazie danych. Takie sytuacje zdarzają się często również w bazach medycznych. Prowadzi to do zbyt optymistycznych wyników. Na razie nie ma dobrego sposobu na uwzględnienie tego typu przypadków.
- Odkryte przez nas reguły są proste i posługują się niewielką liczbą ważnych diagnostycznie testów; w kilku przypadkach udało się odkryć najprostszą znaną strukturę danych. Porównanie z innymi

metodami uczenia maszynowego, w tym z metodami generującymi reguły rozmyte, pokazuje jednoznacznie wyższość proponowanych przez nas metod. Odkryte przez nas reguły te podane są na stronie: <http://www.phys.uni.torun.pl/KIS/projects/rules.htm>

- Nasza działalność w tym zakresie została zauważona przez firmę Bayer Diagnostic z Terrytown, Nowy Jork. Firma ta jest szczególnie zainteresowana opracowaniem systemu wspomaganie diagnostyki dla swoich urządzeń Advia 120 wykonujących testy hematologiczne. Liczymy na szeroki dostęp do danych i rozwój tej współpracy w najbliższych latach.

Publikacje związane z tym tematem:

1. Grudzinski K, Karwowski M, Duch W, Computational Intelligence Study of the Iron Age Glass Data (wysłane)
2. Grudzinski, K. „Similarity-based methods in application to analysis of scientific and medical data”, Praca doktorska, KIS UMK, Toruń 2002

II.2.5. Integracja metod inteligencji obliczeniowej i nowe modele systemów uczących się.

Znanych jest obecnie wiele metod inteligencji obliczeniowej, wyrastających z różnych dziedzin: uczenia maszynowego, sieci neuronowych, algorytmów ewolucyjnych, logiki klasycznej i romytej, rozpoznawania wzorców, statystyki. Każde środowisko ma inne tradycje stosuje własne metody, stąd stosunkowo niewielkie przenikanie się idei opracowanych w pokrewnych dziedzinach. Naszym celem jest integracja metod uczenia maszynowego (ML, machine learning) i statystycznych metod rozpoznawania obiektów z metodami typu sieci neuronowych, opracowanie nowych metod, wynikających z takiego zintegrowanego ujęcia, oraz automatyzacja tworzenia optymalnych modeli do konkretnych zastosowań.

W ostatnich latach opracowaliśmy i rozszerzyliśmy ogólny schemat dla metod opartych na podobieństwie (Similarity-Based Methods, SBM), obejmujący jako szczególne przypadki większość znanych metody rozpoznawania obiektów, takich jak różne warianty k-NN czy kwantyzacji wektorowej LVQ, wiele modeli sieci neuronowych, w tym najbardziej popularne modele MLP i RBF, oraz wiele innych metod. Schemat ten oparty jest na parametryzacji prawdopodobieństw *a posteriori* za pomocą funkcji podobieństwa, sposobu wyboru prototypowych wektorów, sposobu ważenia ich wpływu na obliczane prawdopodobieństwo, liczby wykorzystywanych wektorów prototypowych, sposobu ich wyboru i optymalizacji ich położenia, sposobu wyboru cech do klasyfikacji, sposobu mieszania różnych modeli ze sobą i wielu innych procedur. Opisaliśmy zastosowania tego schematu w zagadnieniach klasyfikacji, aproksymacji, asocjacji i szukania brakujących danych.

Ponieważ nie istnieje jeden systemu, który zawsze najlepiej działa na wszystkich możliwych danych, należy w każdym przypadku szukać najlepiej działającej metody wśród tych, które można wygenerować w schemacie metod opartych na podobieństwie. Korzystając ze schematu SBM rozwinęliśmy **algorytm meta-uczenia**, czyli automatycznego konstruowania optymalnej metody mieszczącej się we wspólnym schemacie, metody najlepiej działającej dla konkretnej bazy danych. Oprócz możliwości wyboru liczby sąsiadów, wyboru i optymalizacji wielu funkcji odległości, można w nim dokonywać selekcji cech i wektorów referencyjnych (prototypów). Optymalizacja miary podobieństwa obejmuje skalowanie odległości i minimalizację funkcji kosztów za pomocą metody symulowanego wyżarzania i metody wielosympleksowej.

W 2002 roku pracowaliśmy nad metodami oceny podobieństwa **wykraczającymi poza paradygmat przestrzeni wektorowych** o ustalonej liczbie wymiarów. Jest to potrzebne do zastosowań bioinformatycznych, chemicznych i innych, w których obiekty mają różną strukturę, trudną do opisania za pomocą ustalonych cech. Pozwoli to na dalszą integrację metod inteligencji obliczeniowej z opartymi na procesach szukania metodami używanymi w sztucznej inteligencji. Wykorzystując programowanie dynamiczne napisano program na znajdowanie transformacji jednego obiektu w drugi w oparciu o ciągi przekształceń o zmiennych kosztach. Niestety w ogólnym przypadku jest to proces obliczeniowo NP-trudny, dlatego skupiono się nad znalezieniem przybliżonych rozwiązań. Schemat SBM idealnie nadaje się do współpracy z tego typu metodami bo nie wymaga reprezentacji w przestrzeni wektorowej, a tylko macierzy podobieństw lub odmienności.

Rozwijaliśmy też metody integracji różnych modeli w oparciu o ideę komitetów modeli. Jest to obecnie duża poddziedzina metod inteligencji obliczeniowej, organizowane są konferencje, numery specjalne a nawet towarzystwa, zajmujące się tylko tym tematem. Celem jest nie tylko poprawa wyników dla złożo-

nych problemów z którymi nie radzi sobie pojedynczy model, ale i zmniejszenie wariacji wyników modelu. W ramach pracy magisterskiej (Ł. Itert) opracowano nową, bardzo obiecującą wersję komitetów „lokalnie kompetentnych modeli”. Testy dały bardzo dobre rezultaty.

Publikacje związane z tym tematem:

1. Duch W, Grudziński K, Meta-learning via search combined with parameter optimization. Intelligent Information Systems, *Advances in Soft Computing*, Physica Verlag (Springer) 2002, pp. 13-22
2. Duch W, Itert L, Competent undemocratic committees. International Conference on Neural Networks and Soft Computing (ICNNSC). *Advances in Soft Computing*, Physica Verlag (Springer) 2002, pp. 406-411
3. Marczak M, Duch W, Grudziński K, Naud A. Transformation Distances, Strings and Identification of DNA Promoters. International Conference on Neural Networks and Soft Computing (ICNNSC), *Advances in Soft Computing*, Physica Verlag (Springer) 2002, pp. 620-625
4. Karol Grudzinski, „Similarity-based methods in application to analysis of scientific and medical data”, Praca doktorska, KIS UMK, Toruń 2002.
5. Łukasz Itert, Komitety systemów klasyfikujących. Praca magisterska, KIS UMK, Toruń 2002

II.2.6. Wizualizacja i redukcja wymiarowości danych.

Wizualizacja służy dwóm celom: wstępnej eksploracji danych i zrozumieniu struktury wielowymiarowych danych dzięki tworzeniu ich wizualnych reprezentacji, oraz próbie zmniejszenia wymiarowości danych. Podejście oparte na minimalizacji miar zachowania topograficznego sąsiedztwa wielowymiarowych danych, czyli skalowanie wielowymiarowe (MDS) oraz odwzorowanie Sammona, daje ciekawsze rezultaty niż samoorganizująca się mapa Kohonena, ale jest bardzo kosztowne obliczeniowo. Opracowano bardzo ciekawe połączenie MDS z metodami hierarchicznej klasteryzacji, jest ono obecnie implementowane i testowane.

W przypadku dużej liczby danych zaproponowano połączenie algorytmu relatywnego mapowania z klasteryzacją; przeprowadzono obliczenia dla bazy danych liczącej ponad 4000 wektorów pokazując, że połączenie takiego podejścia (do klasteryzacji użyto metody LVQ) nie powoduje znaczących różnic w końcowej mapie, pozwalając na wizualizację za pomocą MDS bardzo dużych zbiorów danych. Zbadano metodę mapowania względnego wykrywając interesujące efekty, których do końca nie rozumiemy i obecnie je badamy (nie są to błędy programu).

Podjęto również prace nad wizualizacją rezultatów działania sieci neuronowych, zarówno dla danych statycznych jak i dynamicznych. Wstępne rezultaty znalazły się w ramach pracy magisterskiej i są obecnie przygotowywane do druku. Mapy obrazujące wyjścia sieci neuronowej za pomocą rzutu na rogi wielokąta niosą wiele przydatnych informacji, pozwalając unikać modeli, które są zbyt „pewne siebie” i wybierać optymalne modele spośród kilku o tej samej dokładności. Można też na nich obserwować wpływ regularyzacji i efekty stosowania różnych metod uczenia, zbytniego dopasowania do danych czy zbyt małej złożoności sieci. Publikacja na ten temat została wysłana na początku 2003 roku.

Publikacje związane z tym tematem:

1. Naud A, Duch W, Visualization of large data sets using MDS combined with LVQ. International Conference on Neural Networks and Soft Computing (ICNNSC), *Advances in Soft Computing*, Physica Verlag (Springer) 2002, pp. 632-637
2. Maciej Orłowski, Wizualizacja działania sztucznych sieci neuronowych dla danych statycznych i dynamicznych, Praca magisterska, KIS UMK 2002

II.2.7. Selekcja cech.

Prowadziliśmy intensywne prace nad metodami rankingu cech, selekcji i agregacji dostępnej informacji w celu redukcji wymiarowości. Dokonano przeglądu literatury i opracowano zestaw programów realizujących liczne znane metody. Rozwinięto kilka nowych metod selekcji w oparciu o alternatywne indeksy oceny jakości cech. Metody rankingowe traktują każdą cechę niezależnie, podczas gdy metody selekcji uwzględniają redundancję. Dzięki zastosowaniu metod opartych na technikach haszowania możemy

zastosować każdą ze znanych metod rankingu do selekcji. Przeprowadzono bardzo wiele testów mających na celu określenie wpływu dyskretyzacji na wyniki, ustalenie błędów związanych z dyskretyzacją w oszacowaniu ilości informacji łącznej, warunkowej i wzajemnej, i innych wielkości. Opracowano model sztucznych danych, dla których można jednoznacznie porównać różne metody. Przeprowadzono wiele testów i wykonano obliczenia za pomocą szeregu metod na podzbiorach cech otrzymanych z różnych metod selekcji.

Metody rankingu cech mogą się przydać do skalowania cech. Dr Jankowski opracował algorytm skalowania cech dla metod opartych na podobieństwie; działa ona w oparciu o procedury szukania, początkowo zmieniające wartości współczynników skalujących o 1, potem o 0.5 itd, powtarzając obliczenia w zrandomizowany sposób celem uniknięcia lokalnych minimów; algorytm ten został zaimplementowany w systemie Ghostminer. Przeprowadzono też wstępne badania metod redukcji wymiarowości uzyskując dobre wyniki przy wykorzystaniu skalowania wielowymiarowego.

Publikacje:

1. Duch W, Biesiada J, Winiarski T, Grudziński K, Grąbczewski K, Feature selection based on information theory filters and feature elimination wrapper methods. International Conference on Neural Networks and Soft Computing (ICNNSC), Advances in Soft Computing, Physica Verlag (Springer) 2002, pp. 173-176
2. Duch W, Winiarski T, Grąbczewski K, Biesiada J, Kachel, A, Feature selection based on information theory, consistency and separability indices. International Conference on Neural Information Processing (ICONIP), Vol. IV, pp. 1951-1955, Singapore 2002
3. Jankowski, N. Discrete quasi-gradient features weighting algorithm, International Conference on Neural Networks and Soft Computing (ICNNSC), Advances in Soft Computing, Physica Verlag (Springer) 2002, pp. 194-199

II.2.8. Pozostałe tematy.

Oprócz przedstawionych powyżej głównych tematów w ramach badań własnych podejmowaliśmy też różne drobniejsze tematy związane z metodami inteligencji obliczeniowej.

- Opracowano w ramach pracy magisterskiej (S. Skrzynecki) prostą metodę poprawek *a posteriori* do prawdopodobieństw klasyfikacji, mogącą znaleźć szerokie zastosowanie dla metod opartych na podobieństwie, drzew decyzyj i metod uczenia maszynowego. Metoda została udoskonalona i zastosowana do przypadków wieloklasowych.
- Uwagi na temat wyzwań stojących przed inteligencją obliczeniową są bardzo istotne, mogą bowiem nadać nowy kierunek rozwoju całej dziedzinie. Opisano nowe podejście do budowy sieci neuronowych o różnej skali złożoności zwracając uwagę na rosnącą wiedzę wewnątrz elementu i oddziaływanie między elementami. prowadzi to do sieci, których elementami są prostsze sieci realizujące funkcje traktowane jako lokalne funkcje jednego modułu na wyższym poziomie.
- Pan T. Grzankowski napisał pracę licencjacką traktującą o problemach określania intensywności plamek mikromacierzy; jest to pierwszy krok do uzyskania danych podlegających dalszej obróbce.
- Opracowaliśmy programy pozwalając przekształcać teksty pisane w LaTeXu bezpośrednio na hiperteksty pomocy w systemie Windows.
- Pan S. Soszyński napisał obszerną pracę magisterską na temat komputerów kwantowych i neurokomputerów kwantowych; praca ta powinna się wkrótce ukazać w postaci publikacji książkowej.
- Pan Ryszard Jabczyński napisał pracę magisterską na temat identyfikacji języka za pomocą nagranych próbek mowy.

Publikacje:

1. Duch W, Itert, L, A posteriori corrections to classification methods. International Conference on Neural Networks and Soft Computing (ICNNSC), Advances in Soft Computing, Physica Verlag (Springer) 2002, pp. 412-417
2. Duch W, Mańdziuk J, Quo vadis, computational intelligence? Springer Lecture Notes in Soft Computing 2002, Vol. XX, str. 1-20
3. T. Grzankowski, Analiza danych z mikromacierzy. Praca licencjacka, KIS UMK, Toruń 2002

4. Jankowski N, Grąbczewski, K. From LaTeX to HTML Help, Proceedings of the XIII European TeX Conference, Bachotek, Poland, 2002
5. Sławomir Soszyński, Teoria komputerów kwantowych. Praca magisterska, KIS UMK Toruń 2002
6. Ryszard Jabczyński, Analiza sygnału mowy. Praca magisterska, KIS UMK Toruń 2002

B) Metody komputerowe w fizyce, chemii i naukach o życiu.

Prof. J. Wasilewski z współpracownikami prowadził badania w ramach tematu: „Struktura energetyczna małych cząsteczek”, zajmując się zastosowaniami teorii funkcjonałów gęstości (DFT – Density Functional Theory) do obliczeń struktury energetycznej i widm elektronowo-oscyłacyjnych cząsteczek. Uczestniczył też w badaniach kierowanych przez prof. K. Jankowskiego z Instytutu Fizyki UMK, dotyczących nowych metod opisu efektów korelacji elektronowej w stanach niezamknięto-powłokowych atomów i cząsteczek.

Dr J. Meller prowadził badania nad zastosowaniem metod obliczeniowych typu dynamiki molekularnej oraz metod bioinformatyki rozpoznawania struktur i funkcji białek, metod dopasowania sekwencji w biologii molekularnej i genomice. Do połowy 2001 roku dr Meller pracował w Computer Science Department, Cornell University w USA, na stażu podoktoranckim. Następnie został zatrudniony jako *assistant professor* w Children's Hospital Research Foundation, zajmując się przede wszystkim bioinformatyką.

Prace te prowadzone są głównie w ramach badań własnych. Najważniejsze uzyskane wyniki w tej grupie tematów to:

II.2.9. Teoretyczne badania struktury energetycznej małych cząsteczek.

Kontynuowano badania stanów oscyłacyjnych w małych cząsteczkach metodami wykraczającymi poza elementarne przybliżenie harmoniczne, w szczególności metodami zaburzeniowymi i pseudospektralnymi (DVR – Discrete Variable Representation). Wymieniona tematyka wchodzi m. innymi w zakres pracy doktorskiej mgr S. Zelka *Teoria funkcjonałów gęstości i reprezentacja zmiennej dyskretnej w badaniach stanów oscyłacyjnych cząsteczek*.

1. W oparciu o prace grupy Lighta zbudowano zespół programów komputerowych realizujących podejście Discrete Variable Representation (DVR) do numerycznego wyznaczania poziomów oscyłacyjnych cząsteczek dwu- i trój-atomowych. Kontynuowano badania nad efektywnością stosowania różnych wariantów teorii funkcjonałów gęstości (DFT) w opisie efektów anharmonicznych w widmach oscyłacyjnych cząsteczek trójatomowych, szczególnie na przykładzie stanu podstawowego cząsteczki SO₂. Przedmiotem badań jest także porównanie wyników otrzymywanych metodami numerycznymi i zaburzeniowymi.

Publikacja związana z tym tematem:

S. Zelek, L. Cyrnek, J. Wasilewski: *Lower vibrational states of SO₂. Density Functional and Discrete Variable Representation Approach*. – wysłana do publikacji (THEOCHEM)

Komunikat konferencyjny:

S. Zelek, L. Cyrnek, J. Wasilewski: *Lower vibrational states of SO₂. Density Functional and Discrete Variable Representation Approach*. – Book of Abstracts of International Conference *Current Trends in Theoretical Chemistry IV*, Cracow, 14-15 June 2002:

2. Kontynuowano badania teoretyczne niższych stanów elektronowych glioksalu (C₂O₂H₂) przy zastosowaniu różnych wariantów metody funkcjonałów gęstości (DFT). Do obliczeń energii przejścia elektronowego S₀ – T₁ zastosowano, z dobrym wynikiem, modyfikowany funkcjonał hybrydowy typu B3. Do wyznaczenie powierzchni otwartopowłokowego stanu singletowego zastosowano wariant czasowo zależny teorii (TD DFT); tego typu zastosowania są w literaturze nieliczne. Obliczenia teoretyczne mają na celu interpretację wyników doświadczalnych, otrzymanych przez prof. Heldta w ramach współpracy międzynarodowej najnowszą techniką laserowo indukowanej fluorescencji” (LIF – Laser Induced Fluorescence). Wyniki te wchodzi w zakres pracy doktorskiej mgr S. Zelka, część jest już opublikowana.

Publikacja związana z tym tematem:

1. S. Zelek. *Teoria funkcjonałów gęstości i reprezentacja zmiennej dyskretnej w badaniach stanów oscylacyjnych cząsteczek*. Praca doktorska, UMK, Toruń 2001; obrona odbyła się 6.03.2002 przed Radą Wydziału Chemii UMK.

II.2.10. Opracowanie nowych metod opisu efektów korelacji elektronowej.

Badania te wykonywane były w ramach grantu KBN (kierowanego przez prof. K. Jankowskiego z Instytutu Fizyki UMK), w realizacji którego uczestniczy prof. Wasilewski,

Prowadzone były badania układów modelowych w ramach nowych wersji wieloreferencyjnych metod sprzężonych klasterów (MRCC - Multi-Reference Coupled Clusters) w sformułowaniach walencyjnie-universalnym, stanowo-universalnym i stanowo-specyficznym. Badania dotyczyły wpływu bazy jedno-cząstkowej (formy orbitali molekularnych) na zbieżność rozwinięcia klasterowego dla stanów podstawowych i wzbudzonych, ze szczególnym uwzględnieniem występowania tzw. stanów obcych („intruderów,,). Prof. Wasilewski jest aktywnym uczestnikiem seminarium naukowego dotyczącego teorii korelacji i podstaw teorii funkcjonałów gęstości.

II.2.11. Symulacje komputerowe w fizyce i chemii.

Prowadzono prace rozpoznawcze i studialne w następujących kierunkach: oddziaływanie tlenu molekularnego z atomami i małymi cząsteczkami; zastosowanie różnego typu funkcjonałów mieszanych w metodzie DFT; zastosowanie metod DFT do opisu stanów otwartopowłokowych; zastosowanie reprezentacji zmiennej dyskretnej (DVR - Discrete Variable Representation) do numerycznego wyznaczania poziomów oscylacyjnych małych cząsteczek; jonorodniki aromatyczne jako hipotetyczny składnik „ciemnej materii” w przestrzeni kosmicznej.

Kontynuowano prace nad interpretacją widm elektronowo-oscylacyjnych pirazyny i pirymidyny, uzyskanych metodą laserowo indukowanej fluorescencji (LIF) w międzynarodowym zespole prof. J. Heldta (Uniwersytet Gdański) i prof. C. Ottingera (Max-Planck-Institut fuer Stroemungsforschung w Getyndze). W badaniach zastosowano szereg wariantów teorii funkcjonałów gęstości i rachunku zaburzeń Moellera-Plesseta.

Publikacje związana z tym tematem:

R. Zieliński, *Badanie stanów oscylacyjnych cząsteczki pirazyny w stanach elektronowych: podstawowym ($S_0^1A_g$) i najniższym trypletowym ($T_1^3B_{3u}$)*. – praca magisterska, zakończona i obroniona 11.02, wyniki przygotowywane są do publikacji;.

M. Zduniak *Teoretyczne badania niższych stanów elektronowo-oscylacyjnych cząsteczki pirymidyny*. Praca magisterska w toku, przewidziana do obrony w roku 2003.

II.2.12. Bioinformatyka i nauki o życiu.

W ramach współpracy z jedną z wiodących instytucji amerykańskich w dziedzinie nauk o życiu i ich klinicznych zastosowań, Children's Hospital Medical Center, gdzie dr J Meller pracuje w Department of Pediatric Informatics jako assistant professor, kontynuowano badania w zakresie bioinformatyki i nauk o życiu. Od października 2001 roku bierze w nich udział były doktorant KIS, Rafał Adamczak, trwa procedura powołania prof. Duchy na stanowisko *adjunct professor* w tej instytucji.

W 2002 roku pracowano nad następującymi tematami:

- Przy współpracy z R. Elberem z Cornell University opracowano szereg nowych potencjałów do przewidywania struktury białek i zaimplementowano algorytmy w kolejnych wersjach pakietu obliczeniowego LOOPP (Learning, Observing and Outputting Protein Patterns). Rozpoznaje on struktury białek (fold recognition). Wykorzystano techniki threading, dopasowania sekwencji (sequence alignment), techniki optymalizacji liniowej i analizy statystycznej struktur białkowych oraz tworzenia bibliotek unikalnych fragmentów białek za pomocą porównywania struktura-struktura przy zastosowaniu metod programowania dynamicznego. Potencjały zawierają parametry adaptacyjne a procedura uczenia się uwzględnia istniejące struktury natywne białek i wygenerowane struktury sztuczne, nie istniejące w przyrodzie. Program jest dostępny w postaci serwera sieciowego umożliwiającego automatyczną anotację sekwencji genomycznych. Serwer LOOPP umożliwia automa-

tyczną annotację sekwencji genomowych. Bierze on udział w ocenie metod automatycznych będącą sklasyfikowanym na jednym z czołowych miejsc w kategorii najtrudniejszych do przewidzenia struktur. Można go znaleźć pod adresem:

<http://www.tc.cornell.edu/reports/NIH/resource/CompBiologyTools/loopp/>

- Opracowano nowe, efektywne i dokładne, metody rozpoznawania trzeciorzędowych struktur białkowych, nowe algorytmy porównywania sekwencji aminokwasów (problem jednowymiarowy) oraz nowe algorytmy porównywania topologii białek (problem trójwymiarowy).
- Pracowano nad problemem diagnostyki artretyzmu dziecięcego za pomocą analizy profili mikro-macierzy ekspresji genów z tkanki maziowej, praca została wysłana do druku.
- Przy współpracy z grupą eksperymentalną nowe mechanizmy wiązania polimerazy RNA. Jest to potencjalnie bardzo ważny mechanizm biologiczny.

Publikacje związana z tym tematem:

1. Meller J, Wagner M, Elber R, *Maximum Feasibility Guideline to the Design and Analysis of Protein Folding Potentials*, Journal of Computational Chemistry, 23: 111 - 118 (2002)
2. Meller J, Elber R; *Protein Recognition by Sequence-to-Structure Fitness: Bridging Efficiency and Capacity of Threading Models*, in Computational Methods for Protein Folding: A Special Volume of Advances in Chemical Physics, ed. R. A. Friesner, John Wiley & Sons 2002
3. Scola MP, Thompson SD, Brunner HI, Adamczak R, Meller J, Glass DN, Synovial Tissue Expression Profiles in JRA Onset Types: Analysis of Discriminating Factors by Pattern Recognition Methods, wysłana do Journal of Rheumatology
4. Wagner, M, Meller J., Elber, R. Large-Scale Linear Programming Techniques for the Design of Protein Folding Potentials, Mathematical Programming (w druku)
5. Kuznetsova A. V, Meller J, Schnell, P. O, Nash, J. A, Sanchez, Y, Conaway, J. W, Conaway R.C. and Czyzyk-Krzeska, M.F. VHL binds hyperphosphorylated large subunit of RNA Polymerase II through a proline hydroxylation motif and targets it for ubiquitination. PNAS (w druku)

C) Modelowanie kognitywne (cognitive modeling).

Prace te prowadzimy głównie w ramach badań własnych.

II.2.13. Modelowanie kognitywne.

Kognitywistyka powoli rozwija się w Polsce jako odrębna dyscyplina badawcza. Nasze prace w tym zakresie zmierzają w kierunku empirycznie uzasadnionych teorii umysłu w oparciu o modele sieci neuronowych oraz filozoficzne próby zrozumienia koncepcji umysłu i zagadnień dotyczących świadomości. Teoria powinna nie tylko być pomocną w rozwiązywaniu podstawowych problemów nauk o poznaniu, lecz również prowadzić do nowych konstrukcji systemów uczących się. Zarysy takiej teorii, w postaci hipotez określających prawdopodobne związki pomiędzy zdarzeniami w mózgu i ich korelacji z percepcją, wyższymi czynnościami poznawczymi i treścią świadomości, zostały już nakreślone w kilku pracach i wystąpieniach konferencyjnych. Matematyczne metody opisu zdarzeń mentalnych wiążą się z teorią układów złożonych w zastosowaniu do sieci neuronowych z rekurencją.

Prof. Duch przewodniczy komitetowi naukowemu utworzonego przy współpracy z logikami, filozofami i pedagogami z UMK pisma, „Kognitywistyka i media w edukacji”. Pismo to ma na celu integrację środowiska osób interesujących się kognitywistyką w Polsce, jest jedynym tego typu pismem w Polsce i jest oficjalnym pismem Polskiego Towarzystwa Kognitywistycznego.

W 2002 roku zajmowaliśmy się następującymi zagadnieniami:

- Ukazał się artykuł „Fizyka umysłu” omawiający możliwości modelowania zjawisk mentalnych i połączenia modeli tego typu z neurodynamiką. Artykuł o geometrycznym modelu umysłu jest w druku w Kognitywistyce.
- Ukazał się artykuł o neurokognitywnej teorii świadomości, który prezentuje realistyczne podejście do problemu zrozumienia natury świadomości i pokazuje, jak można stworzyć świadome kompute-

ry, chociaż jest to technicznie bardzo trudne, gdyż nie będą to urządzenia przypominające komputery.

- Artykuł o synchroniczności omawia realistyczne możliwości wyjaśnienia anomalii obserwowanych w eksperymentach prowadzonych od wielu lat w Princeton (zakładając, że nie ma w nich jakiegoś systematycznego błędu).
- Dwa artykuły o tematyce futurologicznej rozważają skutki rozwoju technologii informacyjnych w najbliższych dziesięcioleciach.

Publikacje związane z tym tematem:

1. Duch W, Fizyka umysłu. *Postępy Fizyki 53D* (2002), 92-103
2. Duch W, Neurokognitywna teoria świadomości, *Kognitywistyka i Media w Edukacji*, 2/2001 (5), pp. 47-67 (ukazało się w 2002 roku)
3. Duch W, Synchronicity, Mind and Matter. *International Journal of Transpersonal Studies* 21 (2002) 155-170; Przedrukowane również w: *Neuroquantology*, Vol. 1, January 2003
4. Duch W, Kognitywistyka drogą do zrozumienia człowieka. *Kognitywistyka i Media w Edukacji*, 2/2001 (5), pp. 40-42 (ukazało się w 2002 roku)
5. Duch W (2002) *Sceptycyzm wobec sceptycyzmu (kontynuacja dyskusji o AI)*. *Kognitywistyka i Media w Edukacji*, w druku.
6. Duch W, Future of the information society and information technology. In: *Wissenschaft und Bildung in einer informatorischen Gesellschaft in der Zeit der europäischen Integration*, V Congress of the Societas Humboldtiana Polonorum, „Information Society“, Wrocław 23-24.06.2001, Ed. Ryszard Grzaslewicz, Wyd. Akademii Rolniczej 2002, pp. 1-7.
7. Duch W, Przyszłość technologii informacyjnych i przyszłość książki. *Wirtualna Edukacja* Nr. 9 (2002) (pierwotnie przedstawione na Konferencji na zakończenie programu „Librarius”, Toruń, 19-20.11.2001)

W 2002 roku oprócz wydawania pisma „Kognitywistyka i media w edukacji” opracowano i udostępniono w Internecie notatki do wykładu będącego wstępem do kognitywistyki. Pierwszy semestr obejmował zagadnienia dotyczące filozofii umysłu widzianej z obecnej perspektywy, a drugi głównie zagadnień budowy i sposobu działania mózgu. Materiały te przyłączone są obecnie do wielu stron internetowych, znajdują się pod adresem <http://www.phys.uni.torun.pl/~duch/Wyklady/index.htm>

Udostępniono również wiele tekstów referatów wygłoszonych na konferencjach i na zaproszenie ośrodków polskich i zagranicznych. Można je znaleźć pod adresem:

<http://www.phys.uni.torun.pl/~duch/ref.html>

D) Badania w zakresie dydaktyki wspomaganej metodami komputerowymi.

II.2.14. Badania w zakresie dydaktyki wspomaganej metodami komputerowymi.

W 2002 roku wykonano w ramach badań statutowych następujące prace:

- Prof. J. Wasilewski koordynował prace redakcyjne nad czwartym wydaniem Wydziałowego Informatora ECTS - *Studia na Wydziale Fizyki, Astronomii i Informatyki Stosowanej UMK* w roku akademickim 2002/03, ISBN 83-231-1453-6, str. 396, Wydawnictwo UMK, Toruń, 2002 (skład i łamanie: mgr H. Małek z KIS).
- Prof. J. Wasilewski w ramach działalności jako prodziekan ds. kształcenia Wydziału Fizyki, Astronomii i Informatyki Stosowanej UMK koordynował prace nad powołaniem i wdrożeniem na Wydziale nowego kierunku fizyka techniczna (w tym specjalności technologie informatyczne), oraz wdra-

za-niem nowego typu dwukierunkowego kształcenia nauczycieli – Międzywydziałowego Studium Ma-tematyczno-Przyrodniczego UMK.

- Utrzymujemy i rozszerzamy strony WWW dotyczące wszystkich spraw związanych z komputerami, fizyką komputerową, sieciami neuronowymi, sztuczną inteligencją, kognitywistyką, badaniami nad mózgiem. Ze stron tych korzysta bardzo wiele osób.

Sieci neuronowe, badania nad mózgiem i statystyka:

<http://www.phys.uni.torun.pl/~duch/neural.html> oraz

<http://www.phys.uni.torun.pl/neural/neural.html>

Kognitywistyka <http://www.phys.uni.torun.pl/~duch/cognitive.html>

Sztuczna inteligencja i uczenie maszynowe, <http://www.phys.uni.torun.pl/~duch/ai-ml.html>

Oprogramowanie i bazy danych do testów metod inteligencji obliczeniowej

<http://www.phys.uni.torun.pl/~duch/software.html>

Fizyka komputerowa, <http://www.phys.uni.torun.pl/~duch/fiz-komp/index.html>

- W ramach eksperymentów z nauczaniem na odległość udostępniliśmy wiele stron Internetowych z notatkami do wykładów, pod adresem:

<http://www.phys.uni.torun.pl/~duch/Wyklady/index.htm>

Duch W, Wykłady wstępne o komputerach i oprogramowaniu, Studium Podyplomowe.

Duch W, *Wstęp do kognitywistyki* (wykład monograficzny).

Duch W, *Mózg, umysł i zachowanie* (wykład monograficzny).

Duch W, *Sztuczna inteligencja* (wykład monograficzny).

Duch W, *Inteligencja obliczeniowa* (wykład monograficzny).

- Utrzymujemy strony WWW dotyczące zagadnień związanych z komputerami i oprogramowaniem. Dziedzina ta, zwana w krajach anglojęzycznych „*information science*”, nie wyodrębniła się jeszcze w Polsce.

Adres: <http://www.phys.uni.torun.pl/~duch/book-fsk.html>

Ze stron tych korzysta bardzo wiele osób (około 1000 nowych załadowań strony miesięcznie). Są to pierwsze książki nadające się jako podręcznik do dziedziny określanej w krajach anglosaskich mianem „*information science*”, dotyczącej zastosowań komputerów do przetwarzania informacji, i w niektórych ośrodkach akademickich już znalazły takie zastosowanie.

Ostatnia publikacja związana z tym tematem to:

- M. Berndt-Schreiber, W. Duch, A. B. Kwiatkowska, A. Polewczyński, K. Skowronek (2002) Pokolenie dorastające z komputerem wkracza na uniwersytety - nowe wyzwania edukacyjne. W: *Rola i Miejsce Technologii Informacyjnej w Okresie Reform Edukacyjnych w Polsce*, red. T. Lewowicki, B. Siemieniecki. Wyd. Adam Marszałek, Torun 2002, pp. 307-314.

III. PUBLIKACJE

III.1. W omawianym okresie ukazały się ogółem 33 prace, w tym:

oryginalnych prac badawczych: 15, w tym
 w recenzowanych wydawnictwach międzynarodowych 11,
 w tym z listy Filadelfijskiego Instytutu Informacji Naukowej 5,
 w recenzowanych wydawnictwach krajowych 4;
 oraz 18 doktoratów, prac magisterskich, raportów i prac popularnonaukowych.
 podręczników i skryptów: 0
 publikacji naukowo-popularnych: 6
 redakcji książek naukowych, tłumaczeń: 0
 artykułów, recenzji: 9
 prac doktorskich: 3

Liczba prac opublikowanych w 2002		33
w tym:	Oryginalne prace badawcze	15
	Monografie, podręczniki i skrypty	0
	Publikacje popularno-naukowe i dydaktyczne	6
	Redakcja książek naukowych, tłumaczenia	
	Artykuły, recenzje	12
Liczba prac przyjętych do druku w 2002		3

III.2. Liczba prac przyjętych do druku: 3

Prac wysłanych: 3

III.3. Przykłady publikacji związanych z działalnością naukową KMK:

Podano przy omawianiu poszczególnych zagadnień; pełny spis znajduje się w Załączniku 2b.