PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1121 — 2006
Pozyskiwanie wiedzy i zarządzanie wiedzą

Janina Anna Jakubczyc

Akademia Ekonomiczna we Wrocławiu

IDENTYFIKACJA I ZASTOSOWANIE KONTEKSTU W UCZENIU Z NADZOREM

1. Wstęp

Uwzględnienie kontekstu w bardzo wielu dziedzinach staje się niemal koniecznością, a nie wyborem. Dlatego występuje coraz szersze zainteresowanie wszystkimi problemami związanymi z kontekstem w najprzeróżniejszych dziedzinach życia. Kontekst staje się również nieodłącznym aspektem prawie każdego zadania klasyfikacji. Umożliwia on pełniejszy opis badanego zjawiska, lepsze rozumienie oraz generowanie klasyfikatorów adekwatnych do jego możliwych różnorodności w badanym zjawisku. W artykule zaproponowano sposób identyfikacji kontekstu w zbiorze przykładów uczących i zastosowanie kontekstu w zadaniu klasyfikacyjnym. Struktura artykułu przedstawia się następująco: w punkcie 2 przedstawiono kierunki badań nad kontekstem w uczeniu z nadzorem, część 3 zawiera opis i propozycję rozwiązania problemu, w punkcie 4 opisano sposób uwzględnienia kontekstu, ostatnia zaś część to podsumowanie.

2. Kontekst w uczeniu z nadzorem

Wiele problemów w świecie rzeczywistym jest niemożliwe do rozwiązania bez uwzględnienia ich podstaw, zakresu odniesień, sytuacji i okoliczności wystąpień, a więc bez tego, co nazywane jest kontekstem (por. [Matwin, Kubat 1996]). Problem ten jest szczególnie ważny w obszarze sztucznej inteligencji (por. [Brézillon 2006; Brézillon 2005]). Prace nad kontekstem w tej dziedzinie koncentrują się na poszukiwaniach sposobów jego reprezentacji (por. [McCarthy 1993; Brézillo; Singh i in. 2003; Singh, Lee; Gopalacharyulu i in. 2004]), identyfikacji (por. [Harris i in. 2000; Nurmi, Floréen 2006]), zmienności (por. [Tsymbal 2004; Aha, Goldstone 1992]) oraz jego wpływu na rozwiązywane zadanie (por. [Harris i in. 2000; Lovett, Anderson 1996]).

W obszarze uczenia z nadzorem wielu autorów zwraca uwagę bezpośrednio (por. [Bergadano i in. 1992; Widmer, Kubat 1996] lub pośrednio na obecność różnych kontekstów występujących w rozpatrywanych przez nich problemach maszynowego uczenia (por. [Dent i in.; Holte, Drummont 1994]). Kładą nacisk na potrzebę opracowania algorytmów uczenia umożliwiających ich wykorzystanie w dziedzinach charakteryzujących się zmiennością czasową oraz w dziedzinach kontekstowo-zależnych.

Badania w tym obszarze prowadzone są na trzech płaszczyznach. Pierwsza dotyczy identyfikacji kontekstu w materiale uczącym opisującym jakieś zjawisko (por. [Turney 1996, s. 53-69]). Najczęściej przyjmuje się, że kontekst jest znany, a więc określa się *a priori*, co jest kontekstem (por. [Létourneau i in. 1998]). Z kolei, gdy kontekst nie jest znany, próby jego odkrycia podążają w dwóch kierunkach. Pierwsza zakłada jego zmienność w czasie, a więc konteksty nie są identyfikowane, a tylko rozróżniane poprzez odmienne zachowanie zjawiska lub zmienność opisu pojęć w różnych przedziałach czasu. Badania nad tego rodzaju kontekstami prowadzili G. Widmer i M. Kubat (por. [Widmer, Kubat 1996]). Ich pomysł polega na zastosowaniu uczenia przyrostowego. Przyjmuje się założenie, że kontekst jest nieznany oraz że nie zmienia się zbyt często. Opis pojęcia powstaje stopniowo na podstawie kolejnych przykładów poprzez dopasowywanie wygenerowanego opisu do zachodzących zmian. W procesie tym następuje śledzenie poprawności i kompletności opisów. Wzrastający procent błędnie przeprowadzanej klasyfikacji (rosnący błąd klasyfikacji z okresu na okres) jest wykorzystywany jako heurystyczny sygnał możliwego przesunięcia w kontekście oraz zmian w opisie pojęcia. W takim przypadku wcześniejszy opis jest przechowywany, a generuje się nowy opis. Z reguły przechowuje się kilka opisów tego samego pojęcia – każdy powiązany z innym kontekstem. Koncepcja ta jest bliska wcześniejszemu rozumieniu tzw. dryfu pojęciowego wykorzystywanego w systemie STAGGER, którego autorami są J.C. Schlimmer i R.H. Granger [Granger, Schlimmer 1986].

Drugi kierunek badań nad identyfikacją kontekstu zakłada, że kontekst jest ukryty w opisach rozpatrywanych zjawisk i można go odkryć dzięki identyfikacji atrybutów kontekstowych i kontekstowo zależnych, których definicje podaje P. Turney w pracy [Turney 1993]. Bazują one na analizie rozkładów poszczególnych atrybutów względem atrybutu klasy. Ten przypadek jest przedmiotem zainteresowania autorki. Propozycja szukania kontekstu polega na identyfikacji atrybutów decyzyjnych, kontekstowo zależnych i kontekstowych przy wykorzystaniu algorytmu C4.5 Quilliana bez potrzeby analizy poszczególnych par rozkładów atrybutów.

Na drugiej płaszczyźnie próbuje się znaleźć sposób uwzględnienia kontekstu. Ogólnie można mówić o dwóch sposobach. Pierwszy polega na tworzeniu nowego atrybutu z atrybutu kontekstowego i atrybutu kontekstowo zależnego (strategie normalizacji kontekstowej – por. [Turney 1996]) i dotyczy sytuacji, gdy kontekst jest zawarty w danych. Natomiast druga propozycja rozwiązania polega na zmianie

wartości atrybutu decyzyjnego (dziedziny, gdzie kontekst jest z poza dziedziny) o kontekstową wartość wpływu (kary) (por. [Létourneau i in. 1998]).

Turney w pracy [Turney 1996] proponuje pięć strategii wykorzystania cech kontekstowo zależnych w uczeniu z nadzorem. Trzy pierwsze mówią o tym, że kontekst można usunąć lub dodać. Mianowicie normalizacja kontekstowa ma na celu redukcję wrażliwości cechy kontekstowo zależnej na atrybut kontekstowy. Ideą jest potraktowanie wszystkich atrybutów jednakowo przez usunięcie wpływu kontekstu. Z kolei ważenie kontekstowe to działanie odwrotne do normalizacji kontekstowej. Polega ono na ważeniu atrybutów decyzyjnych, które ma na celu przypisanie im większego znaczenia klasyfikacyjnego w danym kontekście. Jest to więc preferowanie pewnych atrybutów klasyfikacyjych. Kontekstowe rozszerzenie polega na włączeniu do procesu klasyfikacji cechy kontekstowej, a więc traktowaniu jej jako atrybutu decyzyjnego.

Dwie kolejne strategie to rozłączne i w różnej kolejności wykonywane działania na zbiorach atrybutów kontekstowych i niekontekstowych. Wybór klasyfikatora kontekstowego przeprowadza się w dwóch etapach. W pierwszym, klasyfikację przeprowadza się na zbiorze atrybutów kontekstowych, a następnie ten tzw. specjalizowany klasyfikator stosuje się do atrybutów niekontekstowych. Kontekstowe dopasowanie klasyfikacji jest zaś odwróceniem kolejności kroków w selekcji klasyfikatora kontekstowego. Najpierw przeprowadza się klasyfikację na cechach niekontekstowych, a następnie dopasowuje się klasyfikację opartą na atrybutach kontekstowych.

Zdaniem autorki, strategie normalizacji kontekstowej są dość problematyczne, gdyż wniosek płynący z praktycznego wykorzystania sposobów uwzględnienia kontekstu brzmi: „im więcej strategii zastosujesz, tym lepsze uzyskasz wyniki klasyfikacji" (zob. [Turney 1996]). Dlatego preferowanym sposobem jest pośrednie uwzględnienie kontekstu. Polega ono na wytyczeniu granic kontekstów (na podstawie zawartości informacyjnej atrybutu kontekstowego) względem cechy kontekstowej i generowaniu klasyfikatorów dla każdego kontekstu oddzielnie.

Trzecia płaszczyzna obejmuje badania nad zmiennością pojęć poprzez zmianę kontekstu. Znaczenie wielu pojęć może być zależne od pewnych pośrednich kontekstów, a zmiany w kontekstach mogą powodować mniejsze lub większe zmiany w pojęciach. Są one ściśle związane z reprezentacją pojęć, a więc i kontekstu, gdzie kontekst może być reprezentowany przez parametry lub atrybuty kontekstowe (por. [Matwin, Kubat 1996; McCarthy 1993; Nurmi, Floréen 2006]). Ten obszar leży poza tematyką tego artykułu.

3. Identyfikacja kontekstu

Rozważany jest problem identyfikacji i zastosowania kontekstu w uczeniu z nadzorem. Zakłada się, że kontekst znajduje się w nieuporządkowanym (w tym również względem czasu) materiale uczącym. Kontekst, jeśli jest, nie odnosi się do czynników zewnętrznych, nie jest nim także czas. Jak wspomniano w części 2,

podstawą do odkrycia kontekstu są definicje atrybutów podstawowych (decyzyjnych), kontekstowych i kontekstowo zależnych. Przyjmując, że: X_i to atrybut opisujący, Y – atrybut klasy, α_i minimalny, a β_i maksymalny rozmiar kontekstu, p – prawdopodobieństwo, $S_{i,j}$ – zbiór wszystkich atrybutów z wyjątkiem X_i i X_j, $s_{i,j}$ – wartości wszystkich atrybutów w zbiorze $S_{i,j}$, P. Turney w pracy [Turney 1993] przedstawia je następująco:

a) atrybut jest decyzyjny wtedy i tylko wtedy, gdy $\alpha_i = 0$; oznacza to, że istnieją pewne wartości x_i oraz y, dla których $p(X_i = x_i) > 0$, takie, że: $p(Y = y|X_i = x_i) \neq \neq p(Y = y)$;

b) atrybut jest kontekstowy wtedy i tylko wtedy, gdy $\alpha_i > 0$, co oznacza, że dla wszystkich x_i i y zachodzi: $p(Y = y|X_i = x_i) = p(Y = y)$;

c) atrybut X_i jest kontekstowo zależny względem atrybutu X_j wtedy i tylko wtedy, gdy istnieje podzbiór atrybutów $S'_{i,j}$ zbioru $S_{i,j}$, dla których istnieją pewne x_i, x_j, $s'_{i,j}$ oraz y, i dla których nierówność: $p(X_i = x_i, X_j = x_j, S'_{i,j} = s'_{i,j}) > 0$ spełnia następujące warunki: $p(Y = y|X_i = x_i, X_j = x_j, S'_{i,j} = s'_{i,j}) \neq p(Y = y|X_i = x_i, S'_{i,j} = s'_{i,j})$, $p(Y = y|X_i = x_i, X_j = x_j, S'_{i,j} = s'_{i,j}) \neq p(Y = y|X_j = x_j, S'_{i,j} = s'_{i,j})$.

Proponuje się, by identyfikację cech decyzyjnych, kontekstowych i kontekstowo zależnych przeprowadzać, korzystając nie z porównania rozkładów prawdopodobieństw odpowiednich par atrybutów, ale z miary zawartości informacyjnej, a ściślej – z możliwości wyjaśniania atrybutów klasyfikacyjnych w algorytmie drzew decyzyjnych. Sposób postępowania przedstawia się następująco:

a) utworzenie drzewa decyzyjnego na podstawie pełnego materiału uczącego; ten krok umożliwia podział cech na te, które mają bezpośredni wpływ na zadanie klasyfikacji (atrybuty podstawowe), i te, które wpływu nie mają (atrybuty niedecyzyjne);

b) utworzenie drzew decyzyjnych dla wszystkich atrybutów decyzyjnych, przy uwzględnieniu tylko atrybutów niedecyzyjnych; w tym kroku wśród atrybutów niedecyzyjnych szuka się atrybutów kontekstowych, a wśród atrybutów decyzyjnych cech kontekstowo zależnych;

c) analiza par atrybut kontekstowy – atrybut kontekstowo zależny w celu opisania i zastosowania odkrytego kontekstu.

Badania przeprowadzono na danych z repozytorium maszynowego uczenia Credit Approval (zob. [Confidential..]). Zbiór zawiera 656 przykładów opisanych piętnastoma niejawnymi atrybutami, które zostały nazwane kolejnymi liczebnikami: jeden, dwa, ..., piętnaście. Ostatni szesnasty atrybut to decyzja o przyznaniu lub odrzuceniu wniosku kredytowego. Dziewięć atrybutów jest symbolicznych i sześć numerycznych. O wyborze tego przykładu zadecydowała niejawność atrybutów, która uniemożliwia jakiekolwiek sugestie dziedzinowe, np. przy tworzeniu przedziałów dla atrybutów numerycznych oraz duży rozrzut wartości atrybutów.

```
decyzja
dziewięć = t :
        piętnaście = f lub piętnaście = i lub piętnaście = g lub piętnaście = d lub piętnaście = c lub
        piętnaście = h lub piętnaście = b : tak (135.0)
        piętnaście = a :
            czternaście = d : tak (26.0)
            czternaście = f : nie (8.0)
            czternaście = c :
                dwanaście = f : tak (11.0)
                dwanaście = t : nie (13.0)
            czternaście = 14b :
                dziesięć = t : tak (24.0)
                dziesięć = f : nie (27.0)
            czternaście = a :
                sześć = w lub sześć = q lub sześć = m lub sześć = r lub sześć = cc lub sześć = k lub
                sześć = c lub sześć = d lub sześć = x lub sześć = i lub sześć = e lub sześć = aa : tak (56.0)
                sześć = ff : nie (4.0)
                sześć = j : tak (1.0)
dziewięć = f :
        trzy = l lub trzy = p lub trzy = o lub trzy = y lub trzy = r lub trzy = w: nie (91.0)
        trzy = t : tak (2.0)
        trzy = v lub trzy = u lub trzy = s lub trzy = m lub trzy = n lub trzy = z : nie (84.0)
        trzy = k :
            czternaście = c lub czternaście = b lub czternaście = d lub czternaście = a : nie (62.0)
            czternaście = f : tak (8.0)
```

Rys. 1. Drzewo decyzyjne – atrybuty decyzyjne

Źródło: opracowanie własne.

Przygotowanie danych w pierwszym etapie dotyczyło dwóch zadań. Pierwszym było wprowadzenie przedziałów dla atrybutów ciągłych. Wyznacznikiem do ich określenia były ich rozkłady bez odniesienia do opisywanej przez nie klasy. Drugim zadaniem był podział zbioru przykładów na plik uczący i testujący. Oba zbiory były wygenerowane losowo jako zbiory rozłączne w proporcji odpowiednio 70 i 30%.

Przeprowadzone badania przebiegały w kilku etapach. W pierwszym, wygenerowano drzewo decyzyjne[1] dla pełnego zbioru danych. W wyniku otrzymano listę atrybutów decyzyjnych. W celu zapewnienia pełnego oglądu sytuacji klasyfikacyjnej, przy stosunkowo niewielkiej liczbie atrybutów, wyłączono przycinanie drzewa, a zostawiono ograniczenie do minimalnej liczby 10 przypadków.

Atrybutów decyzyjnych było 7, a mianowicie: dziewięć, piętnaście, czternaście, dwanaście, sześć, dziesięć. Wygenerowane drzewo w postaci tekstowej przedstawiono na rys. 1.

[1] Do obliczeń wykorzystano moduł DETREEX pakietu SPHINX firmy AITECH w Katowicach służący do generowania drzew decyzyjnych.

Następnym etapem jest wyjaśnienie każdego atrybutu decyzyjnego tylko atrybutami pozadecyzyjnymi (zgodnie z definicją, atrybut kontekstowy nie wpływa bezpośrednio na przynależność do klasy, ale ma wpływ na zmienną decyzyjną (por. [Turney 1996]). Zadanie to polegało na wygenerowaniu drzew dla każdego atrybutu decyzyjnego (atrybut decyzyjny jako klasa) i wyborze pary: atrybut decyzyjny – atrybut kontekstowy (atrybuty kontekstowe). Wyboru tego dokonano na podstawie trafności klasyfikacyjnej. Zestawienie wyników przedstawiono w tab. 1.

Tabela 1. Zestawienie trafności klasyfikacyjnej dla atrybutów decyzyjnych

Atrybut	Dziewięć	Piętnaście	Trzy	Czternaście	Dwanaście	Dziesięć	Sześć
Trafność klasyfikacyjna	75%	62,78%	30%	31,67%	61,11%	100,0%	38,89%

Źródło: opracowanie własne.

Na podstawie analizy atrybutów decyzyjnych warty rozważenia jest atrybut 'dziewięć', charakteryzujący się najwyższą trafnością klasyfikacyjną. Wyboru atrybutu kontekstowego dokonano, analizując wygenerowane drzewo decyzyjne dla atrybutu decyzyjnego 'dziewięć' zamieszczonego na rys. 2.

```
dziewięć
jedenaście > 2 : t (149.0)
jedenaście <= 2 :
        osiem = d : t (18.0);   osiem = g8 : t (5.0); osiem = a8 : f (222.0); osiem = e8 : t (10.0);
        osiem = f8 : t (14.0)
        osiem = h : t (5.0); osiem = k8 : t (13.0); osiem = i8 : t (3.0); osiem = j8 : t (1.0); osiem = l8
: t (2.0)
        osiem = b :
            siedem = h : t (12.0);        siedem = bb : t (2.0);        siedem = ff : f (1.0);
            siedem = j : f (0.0)
            siedem = z : f (0.0); siedem = o : f (0.0); siedem = dd : f (1.0); siedem = n : f (0.0)
            siedem = v :
                jedenaście <= 0 : f (35.0)
                jedenaście > 0 : t (16.0)
        osiem = c :
            jedenaście <= 0 : f (20.0)
            jedenaście > 0 : t (13.0)
```

Rys. 2. Atrybuty wyjaśniające atrybut decyzyjny 'dziewięć'

Źródło: opracowanie własne.

Na podstawie analizy atrybutów wyjaśniających jedynym możliwym kandydatem na atrybut kontekstowy jest atrybut 'jedenaście'. Ze względu na stosunkowo niewielki wpływ pozostałych atrybutów w prezentowanym drzewie zdecydowano o ich odrzuceniu.

Z kolei atrybut 'dziesięć' jest jednoznacznie wyjaśniony atrybutem 'jedenaście'. Może być wykorzystany zamiennie lub jako sposób na uzupełnienie ewentualnych luk w wartościach tych dwóch atrybutów. Uwzględnienie kontekstu i jego efekty przedstawione zostały w następnym punkcie.

4. Zastosowanie kontekstu

Zidentyfikowana cecha kontekstowa daje możliwość rozróżnienia kontekstów w materiale uczącym i opisania ich za pomocą odrębnych zbiorów reguł (drzew decyzyjnych). Kontekst wyznacza cecha 'jedenaście' i to ona stanowi wyznacznik podziału zbioru uczącego i testującego na grupy o różnych kontekstach. A więc problemem do rozstrzygnięcia jest wyznaczenie przedziałów wartości atrybutu kontekstowego, które by dobrze opisywały zidentyfikowany kontekst. Ponieważ podział wartości atrybutu 'jedenaście' został dokonany automatycznie, postanowiono dokładnie przyjrzeć się wartościom tego atrybutu, korzystając z zawartości informacyjnej względem atrybutu klasy (zob. rys. 3).

Wartości	Zawartość informacyjna
11	0,6363
7	0,5823
5	0,3819
6	0,3637
12	0,3235
0	0,2122
4	0,1623
3	0,0553
2	0,0383
1	0,0026
10	-0,0116
20	-0,0264
8	-0,0264
9	-0,0264
13	-0,0264
14	-0,0264
15	-0,0264

Rys. 3. Zestawienie zawartości informacyjnej dla kontekstowego atrybutu 'jedenaście' względem atrybutu 'dziewięć'

Źródło: opracowanie własne.

Biorąc pod uwagę zawartość informacyjną oraz zastosowanie zasady unikania tworzenia zbyt dużej liczby przedziałów wartości, przyjęto, że satysfakcjonującą liczbą podziału na dwa podzbiory będzie wartość 4. W dalszej analizie uwzględniono oba podziały i rozpatrywano dwa warianty uwzględnienia cechy kontekstowej.

Pierwszy wariant to zbiór przykładów, dla których wartość atrybutu kontekstowego jest mniejsza od 2 równa 2 i zbiór dla wartości większych od 2. W drugim wariancie wartością podziału jest wartość 4. Atrybut kontekstowy stanowił kryterium podziału na pliki o różnych kontekstach, nie był zaś elementem żadnego z tych zbiorów.

Wyniki obu wariantów i klasyfikacji bez uwzględnienia kontekstu przedstawiono w tab. 2.

Tabela 2. Trafność klasyfikacyjna dla dwóch wariantów atrybutu kontekstowego

Trafność klasyfikacyjna bez kontekstu		82,22%
Przedziały wartości	[- ∝; <=2] i [>2; +∝]	[- ∝; <=4] i [>4; +∝]
Trafność klasyfikacyjna	81,36%	87,84%

Źródło: opracowanie własne.

Można zauważyć, że uwzględnienie atrybutu kontekstowego nie musi zwiększyć trafności klasyfikacyjnej, jeśli zbiór jego wartości nie jest poprawnie uporządkowany względem atrybutu klasy (trafność klasyfikacyjna dla kontekstu wyznaczonego przez wartość '2' cechy kontekstowej jest mniejsza od trafności klasyfikacyjnej bez uwzględnienia kontekstu). Niepoprawny podział wartości może nawet, jak w przedstawionym przypadku, obniżyć trafność klasyfikacyjną.

Kwestia, czy zwiększenie trafności klasyfikacyjnej poprzez narzucenie kontekstu na zbiór przykładów uczących o 5,62% faktycznie poprawia jakość klasyfikacji, pozostaje nierozstrzygnięta. Powodem jest nieznajomość kosztów błędnej klasyfikacji oraz to że, gdy kontekst umożliwia właściwą interpretację i opis problemu, może być wystarczającym powodem jego identyfikacji.

5. Podsumowanie

Zaproponowane rozwiązanie wymaga potwierdzenia na większej liczbie przykładów. Takie prace są kontynuowane. Warto zwrócić uwagę na to, iż w tym przypadku zidentyfikowano tylko konteksty wyznaczone przez jeden atrybut kontekstowy. W sytuacjach zaś, w których takich atrybutów będzie więcej, pojawi się problem, jak je na siebie nałożyć oraz jak połączyć wiele kontekstów. Jest to interesujący problem badawczy.

Aplikacja uzyskanych rozwiązań wymaga zastosowania mechanizmu sterowania klasyfikacją nowych przypadków. Parametrem sterującym jest zidentyfikowany atrybut kontekstowy lub ich grupa. Jeśli jest to jeden atrybut to przełączanie między klasyfikatorami można wykonać poprzez wprowadzenie odpowiedniej reguły. Z kolei w przypadku grupy można zaproponować metaklasyfikator.

Identyfikacja atrybutu (atrybutów) kontekstowego służy przynajmniej dwóm celom. Po pierwsze zdecydowanie dokładniejszemu i pełniejszemu opisowi analizowanych zjawisk, co może przyczynić się również do lepszego ich rozumienia. Drugim celem jest zwiększenie efektywności reguł klasyfikacyjnych; jest to naturalna konsekwencja celu pierwszego. W dalszych badaniach stanowić będzie również podstawę do dynamicznej analizy zjawisk poprzez obserwacje zmienności kontekstów.

Literatura

Aha D.W., Goldstone R.L., *Concept Learning and Flexible Weighting*, *In Proceedings of the 14-th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum, Illinois 1992.

Bergadano F., Matwin S., Michalski R.S., Zhang J., *Learning Two-tiered Description of Flexible Concepts*, [w:] *The POSEJDON System*, Machine Learning 1992, s. 5-43.

Brézillon P., Context in Artificial Intelligence: *I. A Survey of the Literature*, http://www-poleia.lip6.fr/~brezil/Pages2/Publications/Index.html, 30.03.2006 r.

Brézillon P., *Context in Artificial Intelligence: II. Key Elements of Contexts*, http://www-poleia.lip6.fr/~brezil/Pages2/Publications/Index.html, 29.09.2005 r.

Brézillon P., *Context in Human Machine Problem Solving: A Survey*. Technical Report 96/29, LAFORIA, 1996, http://:www-poleia.lip6.fr/~brezil/Pages2/Publications/Index.html.

Confidential – quinlan@cs.su.oz.au, Repository of Machine Learning Databases, http://www.ics.uci.edu /~ mlearn/ MLRepository.html]. Credit Approval. Confidential.

Dent L., Bolicario J., McDermott J., Mitchell T., Zabowski D., *A Personal Learning Applications. Proceeding of 1992 Conference on Artificial Intelligence*.

Gopalacharyulu P.V., Lindfors E., Bounsaythip C., Wefelmeyer W., Orešič M., *Ontology Based Data Integration and Context-based Mining for Life Sciences*, W3C Workshop on Semantic Web for Life Sciences, October 2004, Cambridge, MA, USA.

Granger R.H., Schlimmer J.C., *Incremental Learning from Noisy Data. Machine Learning I*, 1986, s. 317-354.

Harries M.B., Sammut C., Horn K., *Extracting Hidden Context*, Kluwer Academic Publishers, Boston.

Harris M.B., Sammut C., Holm K., *Extracting Hidden Context*, Kluwer Academic Publishers, Boston 2000.

Holte R., Drummont C., *Learning Application for Browaing. AAAI Spring Symposium on Software Agculs*, 1994.

Létourneau S., Matwin S., Famili F., *A Normalization Method for Contextual Data: Experience From a Large-scale Application*, Proceedings of the 10th European Conference on Machine Learning (ECML-98). Chemnitz, Allemagne. Du 21 au 24 août 1998. s. 49-54.

Lovett M.C., Anderson J.R., *History of Success and Current Context in Problem Solving: Combined Influences on Operator Selection*, Cognitive Psychology 1996 31(2), s. 168-217.

Matwin S., Kubat M., *The Role of Context in Concept Learning*, Proceedings of the ICML-96, Workshop on Learning in Context, 1996.

McCarthy J., *Notes of Formalizing Context, Proceedings of the Fifth National Conference on Artificial Intelligence*, 1993.

Nurmi P., Floréen P., *Reasoning in Context-aware Systems*, http://www.cs.helsinki.fi/u/ptnurmi/papers/positionpaper, 2006.

Singh S., Vajirkar P., Lee Y., *Context-based Data Mining Using Ontologies*, [w:] *Lecture Notes in Computer Science: Conceptual Modeling*, Springer-Verlag 2003, s. 405-418.

Singh S., Lee Y., *Intelligent Data Mining Framework*, Twelfth International Conference on Information and Knowledge Management (CIKM03), 30.03.2006.

Tsymbal A., *The Problem of Concept Drift: Definitions and Related Work*, http://citeseer.ist.psu.edu/correct/719332, 2004.

Turney P., *The Management of Context-sensitive Features: A Review of Strategies*, Proceedings of the IMCL-96 Workshop on Learning in Context-Sensitive Domains, 1996, s. 53-69.

Turney P.D., *Robust Classification with Context-sensitive Features. In Industrial and Magineering Applications of Artificial Intelligence and Expert Systems*, Scotland Gordon and Breach, Edinburgh 1993.

Widmer G., Kubat M., *Learning in the Presence of Concept Drift and Hidden Context*, Machine Learning 1996.

THE CONTEXT IDENTIFICATION AND APPLICATION IN THE SUPERVISED LEARNING

Summary

The context gives the ability to more adequate description, better understanding and generating more appropriate classifiers for multiply context in studied phenomena. The author has proposed context identification method in the learning example and the way in which this recurring context may be exploited.