



13 ZAGADNIENIA AKTUALNIE PORUSZANE PRZEZ MŁODYCH NAUKOWCÓW

Redaktor Wydania: Krzysztof Piech

AGH Akademia Górnictwo-Hutnicza, Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej, al. A. Mickiewicza 30, Kraków 30-059.

Skład tekstów i projekt graficzny okładki: Marcin Kuczera

Korekty: Krzysztof Piech, Marcin Kuczera

Opracowanie pt. ZAGADNIENIA AKTUALNIE PORUSZANE PRZEZ MŁODYCH NAUKOWCÓW zawiera recenzowane prace naukowe Młodych Naukowców współpracujących z CreativeTime, którzy wzięli udział w Konferencji Młodych Naukowców nt. DOKONANIA NAUKOWE DOKTORANTÓWI – VI edycja – 14.04.2018 w Warszawie, 15.04.2018 w Poznaniu oraz 21.04.2018 w Krakowie. Skład opracowania wykonano na podstawie dostarczonych przez autorów tekstów. Wszystkie artykuły zostały opublikowane na odpowiedzialność ich autorów. Za treść odpowiadają autorzy poszczególnych tekstów.

ISBN: 978-83-63058-83-8

Opracowanie

Niniejsza książka elektroniczna DVD ma służyć młodym naukowcom. Propagujemy podejmowane działania wśród młodych naukowców, wiedzę, innowacyjne badania oraz rozwój nauki. Nauka musi charakteryzować się ciągłym rozwojem. Dzisiejsi naukowcy korzystają z coraz to nowocześniejszych metod badawczych, prowadzą różnego rodzaju projekty, których efekty w nieodległej przyszłości mają służyć całej społeczności i otaczającemu nas środowisku. Niniejsze opracowanie zawiera zbiór zagadnień prezentujących zainteresowania naukowe młodych adeptów nauki.

Młody naukowiec

Absolwenci studiów drugiego stopnia coraz częściej podejmują decyzję o rozpoczęciu studiów doktoranckich. Decyzja ta często podyktowana jest chęcią pozostania na uczelni w charakterze naukowca i wykładowcy. Niestety po otrzymaniu dyplomu doktora nauk tylko część młodych naukowców pozostanie na uczelni macierzystej. Część młodych doktorów zasili inne uczelnie i jednostki naukowe, a zdecydowana większość rozpoczęła kolejny etap swojego życia w instytucjach państwowych i firmach prywatnych. Dlatego też obok realizacji własnych badań naukowych i pisania pracy, doktoranci powinni podjąć wszelkie możliwe działania zmierzające do nawiązania współpracy z firmami prywatnymi, aby realizować dalszą kariery zawodową. Włączanie się doktorantów w różnego rodzaju projekty międzyuczelniane, współpracę w modelu naukowiec-firma, udział we wszelkich konferencjach i szkoleniach o charakterze biznesowo-naukowym zwiększa szansę doktorantów na rozwój naukowy i zawodowy, a przede wszystkim może przynieść upragnioną satysfakcję.

Młodzi naukowcy, którzy pozostali na uczelni wyższej w charakterze często asystenta, adiunkta mają również wiele możliwości nawiązania współpracy ze stale rozwijającym się polskim biznesem. Należy zastanowić się, w jaki sposób przenieść własne dokonania i pomysły naukowe do realizacji w biznesie.

Biznes

Niewątpliwie szansą dla biznesu są innowacje, które niosą ze sobą między innymi młodzi naukowcy. Każdy dobry biznesmen powinien zdać sobie sprawę, że nie ma innowacji bez nowych pomysłów i badań naukowych.

Sami spróbujmy zachęcić właścicieli polskich firm, osoby decyzyjne, menedżerów do nawiązywania współpracy z nami - Młodymi Naukowcami.

Wydawca:

Wydawca: CREATIVETIME, www.creativetime.pl

biuro@creativetime.pl

Skrytka Pocztowa nr 92, 30-093 Kraków 23

Nakład 75 egzemplarzy

Wydanie ISBN

**ZAGADNIENIA AKTUALNIE
PORUSZANE PRZEZ MŁODYCH
NAUKOWCÓW
13**

Wydawca: CREATIVETIME

Kraków 2018

UCZENIE MASZYNOWE I SZTUCZNA INTELIGENCJA JAKO NARZĘDZIA WSPOMAGANIA DECYZJI W ZARZĄDZANIU KAPITAŁEM LUDZKIM ORGANIZACJI

Filip Wójcik

Streszczenie: Kapitał ludzki i intelektualny stanowią podstawę nowoczesnych organizacji opartych na wiedzy. Zarządzanie kompetencjami jest kluczowym aspektem innowacyjności i zapewnienia wysokiego poziomu oferowanych usług. W niniejszym artykule podjęta została próba przedstawienia kompleksowego i automatycznego systemu wspierającego zarządzanie kapitałem ludzkim, w zakresie budowania modeli kompetencji, rozwoju zawodowego i oceny doboru zadań w podmiotach typu *software house*. Algorytm opiera się na kontekstowej analizie semantycznej, połączonej z ekstrakcją fraz kluczowych i procedurami klasyfikacji. Wyniki uzyskane metodą walidacji krzyżowej – 61% precyzji przypisana do wybranych departamentów oraz 75% precyzji doboru zadań wskazują, iż system potrafi rozpoznawać istniejące w strukturze organizacji wzorce. Przedmiotem dalszych badań jest kalibracja systemu w zależności od komórki organizacyjnej oraz przetwarzanie dokumentów aplikacyjnych kandydatów w procesie rekrutacji.

Słowa kluczowe: wspomaganie decyzji, sztuczna inteligencja, zarządzanie, analiza języka naturalnego

1. Wstęp

Rozwiązań z zakresu sztucznej inteligencji kojarzone są zwykle z implementacjami komercyjnymi, wchodzący w skład określonych produktów lub usług – rozpoznawanie głosu i twarzy w smartfonach, systemy grające na giełdzie (tzw. algotrading), czy rekomendacje artykułów w działalności e-commerce. Tymczasem można je z powodzeniem wykorzystać do wspierania bieżącej działalności organizacji, a w szczególności – wspomagania zarządzania kapitałem ludzkim. Nabiera to szczególnego znaczenia w gospodarce opartej na innowacjach – pojawienie się tak zwanych „pracowników wiedzy” bądź profesjonalistów wprowadziło istotne zmiany w klasycznym modelu podziału na „białe kołnierzyki” – osoby zarządzające i podejmujące decyzje, oraz „niebieskie kołnierzyki” – pracowników fizycznych i robotników. Tę nową grupę można scharakteryzować jako składającą się z dobrze wykwalifikowanych specjalistów, którym udzielana jest znaczna swoboda w wykonywaniu zadań kreatywnych, połączona z odpowiedzialnością za rezultaty, generujące większość zysków przedsiębiorstwa [Jemielniak, 2008]. Tacy pracownicy traktowani są ze szczególną attencją, a realizowane przez nich projekty – monitorowane pod kątem zgodności kompetencji zespołu z wymaganiami oraz stopnia ich realizacji.

W tego rodzaju zastosowaniach klasyczne systemy klasy business intelligence – pozwalające na integrację danych z wielu źródeł i ich obróbkę w czasie rzeczywistym (OLAP), nie są wystarczające – złożoność zachodzących interakcji i ich wieloaspektowość skłaniają do wykorzystania innych rozwiązań. Szczególnie istotne dla omawianej problematyki wydają się być tzw. biało-krzynkowe systemy uczenia maszynowego, oparte na regułach i asocjacjach. Posiadają czytelną, zrozumiałą dla odbiorcy końcowego strukturę, dającą wgląd w wewnętrzne mechanizmy, a tym samym – dające się interpretować, w przeciwieństwie do np. sieci neuronowych.

Badania przeprowadzono w dwóch (chcących zachować anonimowość) firmach typu *software house*, tj. zajmujących się wytwarzaniem oprogramowania na zlecenie. Ich działalność, będąca podstawą szeroko rozumianego outsourcingu, oparta jest na właściwej identyfikacji i utylizacji kompetencji [Sobińska, 2015] – zwłaszcza inżynieryjnych w zakresie wytwarzania oprogramowania.

W związku z powyższym, dla celów badań sformułowano następujący problem główny:

W jaki sposób można zastosować algorytmy uczenia maszynowego, w celu wsparcia zarządzania zasobami ludzkimi, w szczególności – kompetencjami w organizacjach opartych na wiedzy.

Opierając się na tak postawionym pytaniu, zdefiniowano problemy szczegółowe:

- Czy można, posługując się algorytmami uczącymi się, zidentyfikować kluczowe kompetencje pracowników i kandydatów?
- Czy można, posługując się tymi samymi algorytmami, skwantyfikować kompetencje posiadane przez pracowników?
- Czy można, w sposób automatyczny, śledzić rozkład kompetencji w organizacji i wizualizować go w sposób czytelny dla odbiorcy?
- Czy można opracować system, pozwalający rekomendować pracownikom zadania, kursy doskonalenia zawodowego i inne aktywności w obszarze zgodnym z ich kompetencjami tak, by kompleksowo wspierać zarządzanie zasobami ludzkimi?

W kolejnych sekcjach zaprezentowano przegląd badań odnoszących się do przedmiotowego problemu bądź problemów pokrewnych, wykorzystane materiały, oraz wyniki własnych badań.

2. Materiały i metody

Znacząca część badań dotyczących wspomagania zarządzania kapitałem ludzkim za pomocą algorytmów, skupia się na problematyce optymalizacji alokacji zadań [Larose i Larose, 2015; Hameed i in., 2017], czasu pracy

i kolejkowania – sprowadzając je w istocie do zagadnień optymalizacyjnych w warunkach ograniczenia zasobów [Onke Hartmann, 1997]. Nie są one ściśle związane z zarządzaniem kapitałem ludzkim *per se*, a raczej traktują go jako formę abstrakcyjnego, matematycznego problemu do rozwiązania. Nie pojawia się tam także zagadnienie kompetencji jako czynnika kluczowego. Istotniejsze ich akcentowanie obecne jest w pracach autorów, starających się zastosować podejście charakterystyczne dla klasycznych systemów ekspertowych. Opierają się one na ustrukturalizowanych ontologiach domenowych, przygotowywanych przez specjalistów w danej dziedzinie [Ley i in., 2008]. Zaproponowano szereg zabiegów standaryzacyjnych, mających ujednolicić semantykę modeli kompetencyjnych i sposób ich budowania [Sampson, Karampiperis i Fytros, 2007]. Inni autorzy [Naeve, Sicilia i Lytras, 2008] sugerowali także ściśle powiązanie modelowania kompetencji z rozwojem zawodowym i ciągłym kształceniem pracowników. Schemat wnioskowania i systemu quasi-uczącego się, omówiony w publikacji z 2012r. [García-Barriocanal, Sicilia i Sánchez-Alonso, 2012], stanowił próbę zintegrowania opisywanych pojęć. Oparty był jednakże na manualnej definicji kompetencji oraz stopnia jej dopasowania, wprowadzonej do systemu przez eksperta.

W toku badań nad algorytmami prezentowanymi w niniejszym artykule, wykorzystano szereg istniejących podejść, kombinując je ze sobą i modyfikując. Pierwszym i najważniejszym jest zestaw algorytmów znanych pod nazwą analizy asocjacyjnej (ang. *Association analysis*), a w szczególności – tzw. *A Priori* [Borgelt i Kruse, 2003], pozwalający wyszukiwać prawdopodobieństwa warunkowe współwystępujących elementów (tzw. zbiorów częstych). Naturalnym jego rozszerzeniem, stosowanym podczas klasyfikacji obiektów, jest procedura CBA [Liu, Hsu i Ma, 1998] (ang. *Classification based on association rules*) pozwalający konstruować reguły w postaci warunkowej: jeśli poprzednik to następnik, gdzie zarówno poprzednik jak i następnik są zbiorami częstymi elementów. Metody ekstrakcji wyrażeń i identyfikacji fraz kluczowych zapożyczono z klasycznej analizy języka naturalnego [Manning i Schütze, 1999] (ang. *Natural language processing NLP*), a w szczególności procedury takie jak stemming, lematyzacja i konstrukcja n-gramów (tj. złączeń współwystępujących n wyrażeń). Należy podkreślić, iż językiem dominującym w analizowanych podmiotach był angielski. Kolejnym z zastosowanych algorytmów był zmodyfikowany system rekommendacyjny zwany filtracją kolaboratywną (ang. *collaborative filtering*), wykorzystywaną w podmiotach komercyjnych i branży e-commerce do sugerowania kolejnych zakupów [Sarwar i in., 2001; Schafer i in., 2007].

Do klasyfikacji wykorzystano po jednym, reprezentatywnym algorytmie z każdej kategorii – białoskrzynkowych (czytelnych dla człowieka) i czarnoskrzynkowych (tj. posługujących się abstrakcjami matematycznymi niezrozumiałymi dla odbiorcy), mianowicie: drzewa decyzyjne C4.5 i sieci neuronowe (ang. *multi layer perceptron*). Oba algorytmy uzyskują bardzo dobre wyniki w zbliżonych zastosowaniach [Cichosz, 2007; Hastie, Tibshirani i Friedman, 2009].

3. Wyniki

Jak wspomniano w sekcjach wcześniejszych, do badań wybrano dwa podmioty typu *software house* – przedsiębiorstwa wykonujące oprogramowanie na zlecenie klientów lub działające w ramach szeroko rozumianego *outsourcingu* usług IT. W tego rodzaju firmach pojęcie kompetencji jest rozumiane dość wąsko i w praktyce sprowadza się do znajomości określonych technologii lub rozwiązań. Co więcej, podmioty te w swojej praktyce zarządzania projektami posługują sięnaną metodologią *agile*, wspomaganą komputerowymi metodami rozliczania zadań i czasu pracy. Takie warunki ułatwiły zebranie niezbędnych informacji oraz charakterystyk kandydatów, pracowników, departamentów oraz samych zadań.

Zgodnie ze sformułowanym celem badań, przystąpiono do próby zaadresowania zidentyfikowanych wcześniej problemów. **W tym celu sformułowano i poddano falsyfikacji następujące hipotezy badawcze:**

- Można, za pomocą algorytmów analizy semantycznej oraz kontekstowej języka naturalnego, zidentyfikować kluczowe kompetencje pracowników organizacji oraz kandydatów.
- Kompetencje odnalezione w punkcie pierwszym, poddają się kwantyfikacji za pomocą algorytmu badania częstości występowania w zbiorze oraz względnego ładunku informacyjnego.
- Znając numeryczny rozkład kompetencji w organizacji, można przedstawić je za pomocą grafów, obrazujących rozkład w poszczególnych departamentach i wydziałach.
- Można, posługując się algorytmami analizującymi podobieństwo wektorów, oraz klasyfikatorami zbudować system rekommendacyjny zadań.

W kolejnych sekcjach przedstawione zostaną wyniki częściowe dla każdej z hipotez oraz ogólne konkluzje po zakończeniu procedury.

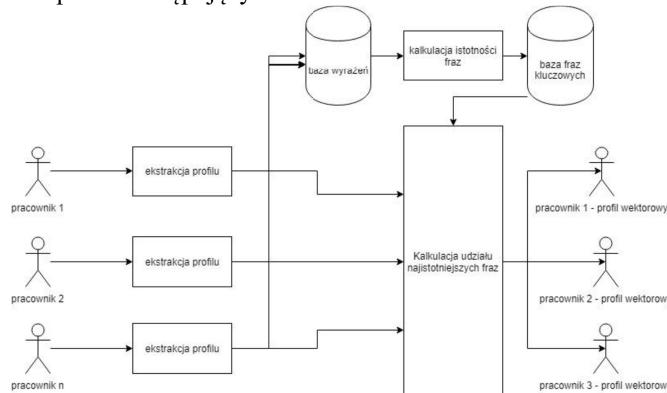
Semantyczna identyfikacja kompetencji

W celu identyfikacji kompetencji pracowników jednej z organizacji zastosowano algorytmy analizy semantycznej i ekstrakcji fraz kluczowych. W tym celu wykorzystano:

- Wyciąg historii wykonywanych przez pracownika zadań, z których każde oznaczone było tzw. „tagami biznesowymi”, tj. skrótownymi opisami najważniejszych funkcji i elementów.
- Nazwę aktualnie zajmowanego oraz wszystkich poprzednio piastowanych stanowisk (przechowywaną w systemie zarządzania kadrami)

- Opis własny pracownika z wewnętrznego korporacyjnego portalu społecznościowego.

Proces można przedstawić w sposób następujący:



Ryc.1: Tworzenie profili semantycznych pracowników. Źródło własne

- Dla każdego pracownika pobrano komplet informacji, zgodnie z przedstawionym wcześniej opisem
- Opisy wprowadzono do wspólnej bazy
- Na bazie wykonano szereg operacji analizy semantycznej [Manning i Schütze, 1999]:
 - Proces usuwania tzw. słów-stopów (ang. *stopwords*), czyli wyrażeń nieniosących żadnej realnej wartości informacyjnej oraz stemming i lematyzacja, czyli znajdowanie ich form uproszczonych
 - Proces kalkulacji istotności każdego słowa Tfifd (ang. *term-frequency inverse document frequency*), [Manning i Raghavan, 2009]. Miarę tę można interpretować jako wagę słowa, rosnącą w miarę pojawiania się w pojedynczym dokumencie, ale malejącą w miarę powszechnego występowania w całej bazie tekstów. Tfifd promuje wyrażenia rzadkie, charakterystyczne dla pojedynczych wystąpień – w tym kontekście charakterystyczne cechy pracowników.
- Wyliczono efektywną wagę kompetencji w zależności od poziomu danej osoby (jeśli takie informacje były dostępne), opierając się na informacjach kontekstowych. W tym celu wykorzystano procedury tzw. zagnieżdżania słów (ang. *word embeddings*) – dla każdego uporządkowanego wektora wyrażeń wyznacza się obszar oddziaływanego, w którym sprawdzane jest współwystępowania innych słów-kluczy [Levy i Goldberg, 2014]. Rozmiar kontekstu w badaniu określono empirycznie (taki, dla którego algorytmy uzyskiwały najlepsze rezultaty). Mnożniki „siły kompetencji” określono, kierując się praktyką biznesową w analizowanych podmiotach. I tak:
 - Mnożnik x2 dla osób wyjątkowo doświadczonych: *senior, expert, specialist*
 - Mnożnik x1 dla osób o przeciętnym/normalnym doświadczeniu: *regular*
 - Mnożnik x0.5 dla osób z niewielkim doświadczeniem: *junior, student, trainee*
 I tak, przykładowo, *Senior Java developer*, przy założeniu, iż słowo-klucz *Java* posiada wagę tfifd 0.86, uzyskałby efektywną wagę 1.72, natomiast *Junior Java developer*, otrzymałby wagę 0.43.
- Po otrzymaniu wektorów tfifd dla każdej z kompetencji, przekształcono do postaci macierzowej profile poszczególnych pracowników.

Wizualizacja rozkładu kompetencji

Mając dane na temat posiadania przez pracowników określonych kompetencji o określonym natężeniu/wadze, połączono te informacje z identyfikatorami departamentów i zespołów, do których należą. W tym celu wykorzystano wewnętrzne dane organizacji, pochodzące z systemu zarządzania personelem.

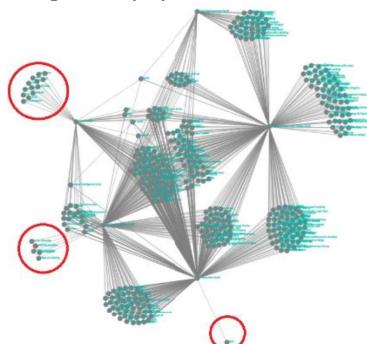
Wykorzystując tak otrzymaną strukturę danych, przeprowadzono analizę asocjacyjną, próbując wyznaczyć najczęściej współwystępujące grupy elementów: Departament – kompetencja, Zespół-kompetencja oraz Grupa-kompetencja. Posłużyono się w tym celu algorytmem *A priori* [Borgelt i Kruse, 2003] - wyszukuje on tylko częste (wg. zdefiniowanego progu) elementy, pary elementów, trójkę, itc. Znalezione elementy częste służą do skonstruowania tzw. reguł asocjacyjnych, będących w istocie regułami decyzyjnymi w formie:

JEŚLI poprzednik TO następnik

gdzie zarówno poprzednik jak i następnik, to elementy częste, pochodzące z pierwszej fazy działania algorytmu. Reguły opisywane są szeregiem własności, pozwalających ustalić ich istotność i przydatność w analizie [Lesh, Zaki i Ogihara, 1999; Morzy, 2013].

Reguły opisujące częstość występowania kompetencji w departamentach/wydziałach zwizualizowano, formując „grały kompetencji”, zbliżone koncepcyjnie do ontologii, używanych np. w systemach wizualizacji wiedzy ekonomicznej [Dudycz, 2013]. W rezultacie otrzymano graf skierowany, w którym wierzchołkami

początkowymi są kompetencje, wagami krawędzi – prawdopodobieństwa warunkowe ich wystąpienia i średnie „natężenie” kompetencji (zgodnie z wcześniejszą definicją), a wierzchołkami docelowymi – departamenty/wydziały. Przykład przedstawia poniższy rysunek:



Ryc.2: Asocjacyjna mapa kompetencji organizacji z zaznaczonymi obszarami anomalii. Źródło własne

W toku analizy przedmiotowego zbioru danych dla jednego z podmiotów uzyskano następujące konkluzje biznesowe:

- Departament testów, posiadający szerokie kompetencje w zakresie jakości oprogramowania (ok. 85% prawdopodobieństwo warunkowe), nie posiada prawie żadnych umiejętności z zakresu Javy (ok. 1% prawdopodobieństwa). Departament liczył ponad 100 osób.
- Natężenie kompetencji w Javie miało strukturę 80% seniorów / 15% przeciętnych / 5% juniorów. W zakresie testowania oprogramowania: 65% seniorów / 30% przeciętnych / 5% juniorów.

Konkluzją biznesową z tak przeprowadzonej analizy było a) stwierdzenie, iż występuje – znaczna przewaga programistów Javy nad testerami w tej technologii b) niewielki odsetek programistów zna się na testowaniu własnej technologii. Biznesowym rezultatem było: a) zintensyfikowanie szkoleń w zakresie testów wśród programistów Javy b) zintensyfikowanie szkolenia testerów w zakresie Javy, c) zatrudnienie większej ilości testerów z dłuższym stażem.

Klasyfikacja pracowników na podstawie kompetencji i system doboru zadań

Kompetencje skwantyfikowane i opisane w sposób przedstawiony w punkcie 3.1 posłużyły do zbudowania klasyfikatora, wykorzystującego je jako zmienne niezależne (X) i przypisującego je do dwóch rodzajów zmiennych zależnych – departamentu/wydziału lub też konkretnego zadania. Do szkolenia algorytmów wykorzystano metodologię walidacji krzyżowej ([Cichosz, 2007; Hastie, Tibshirani i Friedman, 2009; Morzy, 2013; Zaki i Meira, 2014]) polegającej na k-krotnym dzieleniu zbioru już istniejących danych na część treningową i walidacyjną. Systemy uczone były na danych treningowych, a testowane na walidacyjnych. Następnie przeprowadzono testy istotności statystycznej dla różnic pomiędzy trafnością dla każdego z modeli.

- Przypisania do departamentów (7 możliwych) badano na próbce 4019 pracowników, u których ekstrakcja semantyczna zidentyfikowała 151 kluczowych cech
- Rekomendacje zadań (1953 możliwe) testowano na próbce 50 pracowników z jednego zespołu, u których ekstrakcja semantyczna zidentyfikowała 151 kluczowych cech

Uzyskane wyniki przedstawia tabela poniżej (wartości uśrednione ze względu na ograniczone miejsce):

Tab.1: Wyniki klasyfikacji do departamentów. Źródło własne

Model	Precyzja		Czułość	
	Średnia	Odchylenie std.	Średnia	Odchylenie std.
Sieć neuronowa	61%	3%	56%	6%
Las losowy	55%	2%	49%	2%
XGboost	54%	4%	50%	1.5%

Tab.2: Wyniki rekomendacji zadań. Źródło: własne

Model	Precyzja		Czułość	
	Średnia	Odchylenie std.	Średnia	Odchylenie std.
Sieć neuronowa	52%	9%	49%	12%
Las losowy	58%	2%	51%	1%
XGboost	56%	1.5%	52%	2%
System rekomendacyjny	75%	5%	70%	3%

Przedstawione wyżej wyniki dowodzą, iż:

- Modele sieci neuronowych sprawdzają się lepiej w stosunku do innych przy klasyfikacji osób do departamentów/wydziałów Firmy
- System rekomendacyjny filtracji kolaboratywnej w najlepszym stopniu rekomenduje pojedyncze zadania dla pracowników
- W przypadku obu zadań, trafność predykcji przekracza 50% (wartość bazowa), aczkolwiek przypisywanie osób do departamentów oscyluje wokół tej granicy. Rozbicie predykci najlepszego klasyfikatora (sieci neuronowej) na trafność i czułość w ramach pojedynczego departamentu ujawnia, iż wyniki znacząco różnią się w zależności od komórki organizacyjnej:

Tab.3: Wyniki sieci neuronowej w rozbiciu na departament. Źródło własne

Departament	Precyzja	Czułość	Ilość obserwacji
1	74%	27%	473
2	40%	73%	817
2	54%	40%	442
3	69%	26%	514
4	63%	36%	342
5	100%	100%	329
6	88%	42%	500
7	35%	60%	602
ŚREDNIA:	65%	51%	

Przyczyną takiego stanu rzeczy jest fakt, iż niektóre departamenty odznaczają się mniej wyraźną charakterystyką kompetencji niż pozostałe. Oznacza to, iż system klasyfikacyjny musi być optymalizowany dla pojedynczych wydziałów oddziennie, nie może być stosowany globalnie dla całej Firmy.

4. Dyskusja

Przedstawione w części wstępnej systemy klasyfikacji kompetencji – czy to oparte na ontologiach, systemach ekspertowych, czy to na numerycznej optymalizacji zadań, nie rozwiązyują całkowicie problemu ekstrakcji, przetwarzania i analizy kompetencji. Algorytm oparty na analizie semantycznej i reprezentacji wektorowej zapewnia jednolite ujmowanie zagadnienia, bez względu na cel końcowy – czy to wizualizację rozkładu, czy też przypisywanie osób do departamentów lub rekomendację zadań. Co więcej, w obecnej postaci nie wymaga ingerencji człowieka poza wskazaniem podstawowych parametrów dla modeli oraz słów-stopów. Odróżnia go to od np. systemów opartych na ontologiach czy mapach pojęć [García-Barriocanal, Sicilia i Sánchez-Alonso, 2012; Dudycz, 2013].

Nie oznacza to, iż system ten jest uniwersalny, czy też najlepszy dla opisywanych zadań. Grupa twierdzeń pod wspólną nazwą *no free lunch theorem* [Hastie, Tibshirani i Friedman, 2009] wyraźnie wskazuje, iż złożoność obliczeniowa zadań z zakresu uczenia maszynowego wyklucza istnienie tego rodzaju optymalnych procedur. Autonomiczny charakter (samodzielna ekstrakcja cech kluczowych i optymalizacja parametrów modelu) oszczędza jednak czas i nakład ludzkiej pracy. Niemniej jednak, jak pokazały badania, trafność predykci waha się znacznie w zależności od departamentu, co może wykluczać jego zastosowanie w mało wyróżniających się środowiskach.

W chwili pisania niniejszego opracowania, hipotezy odnoszące się do jednego z obranych celów badawczych wciąż są w trakcie weryfikacji – dotyczy to oceny kandydatów. Ze względu na trudności z integracją życiorysów w rozmaitych formatach (.doc, pdf, LaTeX, OpenOffice, etc.) wciąż trwa szkolenie osobnego algorytmu ekstrakcji semantycznej, opartego na sieciach neuronowych do rozpoznawania pisma i obrazu (OCR).

5. Wnioski

Przeprowadzone dotąd badania pozwoliły na zweryfikowanie wszystkich hipotez badawczych w zakresie odnoszącym się do osób aktualnie zatrudnionych w badanych organizacjach. Udało się zatem potwierdzić, iż za pomocą analizy semantycznej można zidentyfikować kluczowe kompetencje pracowników, a następnie skwantyfikować je pod kątem natężenia w poszczególnych wydziałach. Na tej podstawie uzyskano wizualizację grafową, będącą przedmiotem osobnej analizy, prowadzącej do decyzji biznesowych w zakresie zarządzania kapitałem ludzkim. W oparciu o wektorową reprezentację kompetencji pracowników, skonstruowano systemy klasyfikacyjne – przypisujące osoby do najbliższych im departamentów (średnia precyzja 61%, średnia czułość 56%), oraz rekomendacyjny – wskazujący najlepiej dobrane zadania (średnia precyzja 75%, średnia czułość 70%). W przypadku przypisania do departamentów wariancja czułości była bardzo duża ze względu na zróżnicowanie jednostek organizacyjnych. Ten aspekt będzie przedmiotem dalszych badań i optymalizacji.

W toku znajdują się badania i eksperymenty dotyczące automatycznej ekstrakcji kompetencji z dokumentów aplikacyjnych kandydatów. W tym celu wykorzystany będzie system rozpoznawania tekstu na obrazach (konwolucyjna sieć neuronowa), a następnie system analizy wykorzystujący sieci neuronowe o pamięci zmiennej (ang. *Long-short term memory neural network*).

6. Literatura

- Borgelt, C. i Kruse, R.** 2003. Induction of Association Rules: Apriori Implementation. Härdle W., Rönz B. (eds) Compstat. Physica.
- Cichosz, P.** 2007. Systemy uczące się. Wyd. 2. Warszawa: Wydawnictwa Naukowo-Techniczne.
- Dudycz, H.** 2013. Mapa pojęć jako wizualna reprezentacja wiedzy ekonomicznej. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego.
- García-Barriocanal, E., Sicilia, M.-A. i Sánchez-Alonso, S.** 2012. Computing with competencies: Modelling organizational capacities, Expert Systems with Applications, 39(16), s. 12310–12318.
- Hameed, M., Khalid H., Qamar U.** 2017. Optimizing software project management staffing and work-force deployment processes using swarm intelligence, w 2017 Computing Conference. IEEE, s. 78–84.
- Hastie, T., Tibshirani, R. i Friedman, J.** 2009. The Elements of Statistical Learning. 2. wyd. Springer.
- Jemielniak, D.** 2008. Praca oparta na wiedzy: praca w przedsiębiorstwach wiedzy na przykładzie organizacji high-tech. Warszawa: Wydawnictwa Akademickie i Profesjonalne.
- Joel-Edgar, S. i Gopsill, J.** 2018. Understanding user requirements in context: A case study of developing a visualisation tool to map skills in an engineering organisation. W 2018 International Conference on Information Management and Processing (ICIMP). IEEE, s. 6–10.
- Larose, D. T. i Larose, C. D.** 2015. Data mining and predictive analytics. Wiley-Blackwell 2 ed.
- Lesh, N., Zaki, M. J. i Ogihara, M.** 1999. Mining features for sequence classification. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '99. New York, New York, USA: ACM Press, s. 342–346.
- Levy, O. i Goldberg, Y.** 2014. Dependency-Based Word Embeddings. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), s. 302–308.
- Ley, T.** 2008. Modeling competencies for supporting work-integrated learning in knowledge work. Journal of knowledge management. Emerald Group Publishing Limited, 12(6), s. 31–47.
- Liu, B., Hsu, W. i Ma, Y.** 1998. Integrating Classification and Association Rule Mining. KDD'98 Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. New York: AAAI Press, s. 80–86.
- Manning, C. D. i Raghavan, P.** 2009. An Introduction to Information Retrieval, w Online, s. 1.
- Manning, C. D. i Schütze, H.** 1999. Foundations of statistical natural language processing. Cambridge, 2 wyd. Massechussets: MIT Press.
- Morzy, T.** 2013. Eksploracja danych: metody i algorytmy. Warszawa: Wydawnictwo Naukowe PWN.
- Naeve, A., Sicilia, M.-A. i Lytras, M. D.** 2008. Learning Processes and processing learning: from organizational needs to learning designs. Journal of Knowledge Management. Emerald Group Publishing Limited, 12(6), s. 5–14.
- Onke Hartmann, S.** 1997. A Competitive Genetic Algorithm for Resource-Constrained Project Scheduling, Naval Research Logistics, 45, s. 733–750.
- Sampson, D., Karampiperis, P. i Fytros, D.** 2007. Developing a common metadata model for competencies description. Interactive Learning Environments. Taylor & Francis, 15(2), s. 137–150.
- Sarwar, B. i in.** 2001. Item-based collaborative filtering recommendation algorithms. Proceedings of the tenth international conference on World Wide Web - WWW '01, s. 285–295.
- Schaeffer, S. E.** 2007. Graph clustering. Computer Science Review, 1(1), s. 27–64.
- Schafer, J. i in.** 2007. Collaborative Filtering Recommender Systems. The Adaptive Web, s. 291–324.
- Sobińska, M.** 2015. Przewodnik sourcingu IT. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego.
- Yonamine, T. i in.** 2015. A means for visualization of skills in software development. 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE). IEEE, s. 1–5.
- Zaki, M. J. i Meira, W.** 2014. Data mining and analysis : fundamental concepts and algorithms. Cambridge: Cambridge U.P.

Nazwa instytucji: Uniwersytet Ekonomiczny we Wrocławiu, Wydział Zarządzania, Informatyki i Ekonomii, Katedra Inteligencji Biznesowej w Zarządzaniu

Opiekun naukowy: dr hab. Iwona Chomiak-Orsa, prof. UE

Adres do korespondencji: filip.wojcik@ue.wroc.pl