

Popularnonaukowe streszczenie projektu

Entropia jest to powszechnie wykorzystywana w nauce miara nieuporządkowania. Im większą dany system posiada entropię, tym bardziej jest nieuporządkowany, lub, innymi słowy, tym większa panuje niepewność odnośnie stanu tego systemu. Weźmy jako przykład pudło wypełnione białymi i czarnymi kulkami. Jeśli kulek każdego koloru jest tyle samo, to istnieje największa możliwa niepewność odnośnie koloru wylosowanej na ślepo kulki, zatem entropia takiego pudła jest maksymalna. Jeśli pudło byłoby uporządkowane, tj. zawierało kulki tylko jednego koloru, to entropia takiego pudła wynosiłaby 0 i nie istniałaby żadna niepewność odnośnie wyniku losowania kulki. Tę prostą analogię można rozciągnąć z łatwością na dowolny system złożony, w ramach którego zachodzi proces jakiegokolwiek uczenia.

W świecie sieci złożonych algorytmy przetwarzania tych sieci często stoją przed koniecznością oszacowania różnych cech poszczególnych wierzchołków. Sieci charakteryzujące się regularnością posiadają niewielką entropię, przy czym niekoniecznie musi to oznaczać równomierny rozkład np. stopni wierzchołków, ich bliskości czy pośrednictwa. Może to np. oznaczać, że w sieci występuje kilka zbiorów wierzchołków, które mają ściśle określone wartości tych miar centralności, z niewielką wariancją. Wiedza o entropii poszczególnych miar centralności może być bardzo przydatna w algorytmach przetwarzania sieci, ponieważ pozwala sterować algorytmami uczenia kierując je w kierunku rosnącej lub malejącej entropii. Warto przy tym wspomnieć, że istnieje wiele definicji entropii, które próbują uchwycić różne aspekty nieuporządkowania sieci złożonej, i definicje te niekoniecznie są nawzajem spójne. Na dodatek wiele z definicji entropii zaproponowanych wcześniej w literaturze wymaga przeprowadzenia bardzo kosztownych i czasochłonnych obliczeń, co w obliczu rozmiarów współczesnych sieci złożonych (np. sieci społecznościowych) wyklucza możliwość praktycznego zastosowania takich definicji.

Celem niniejszego projektu jest zbadanie cech poszczególnych entropii w świecie sieci złożonych, ze szczególnym naciskiem położonym na sieci ze znakiem (ang. *signed networks*), sieci wielowarstwowe (ang. *multiplex networks*) oraz sieci wielomodalne (ang. *multimodal networks*). Sieci ze znakiem to sieci, w których krawędzie między poszczególnymi wierzchołkami posiadają wartościowość (najczęściej pozytywną i negatywną) reprezentującą różne semantyki krawędzi (przykładowo, sympatię i niechęć). W sieciach wielowarstwowych wierzchołki mogą być jednocześnie połączone krawędziami reprezentującymi zupełnie różne rodzaje związków (przykładowo, jeden rodzaj krawędzi reprezentuje przepływ poczty elektronicznej a drugi rodzaj krawędzi reprezentuje stosunki służbowe między osobami przedstawionymi w postaci wierzchołków). Wreszcie, w sieciach wielomodalnych wierzchołki należą do wielu nawzajem rozłącznych zbiorów reprezentujących klasy obiektów (przykładowo, osoby i firmy, wraz z ich wzajemnymi powiązaniem). Głównym celem projektu jest zatem opracowanie intuicyjnych, obliczeniowo "tanich" definicji entropii i rozciągnięcie tych definicji na nowe klasy sieci złożonych, a także dowiedzenie użyteczności zaproponowanych miar poprzez ich wykorzystanie w algorytmach przetwarzania sieci.

Reasumując, w pierwszej fazie projektu nastąpi adaptacja różnych definicji entropii do świata sieci złożonych oraz uproszczenie tych miar w taki sposób, aby były obliczeniowo aplikowalne do współczesnych sieci złożonych. W drugiej fazie projektu nastąpi operacjonalizacja definicji zaproponowanych w fazie pierwszej, tzn. włączenie opracowanych definicji entropii do wybranych klas algorytmów przetwarzania sieci. Wybraliśmy na tym etapie trzy rodzaje algorytmów: klasyfikację sieci (tzn. przypisywanie sieci do właściwej generalnej klasy sieci), modularyzację sieci (poszukiwanie grup wierzchołków tworzących spójne podgrupy w sieci), oraz kompresję sieci (poprzez łączenie wierzchołków lub poprzez jej próbkowanie). Tym samym mamy nadzieję dowieść uniwersalnej użyteczności pojęcia entropii w przetwarzaniu sieci złożonych.