

Integracja danych z wielu sensorów

Sławomir Gilewski

Dariusz Max Adamski

Streszczenie—Integracja danych z wielu sensorów to bardzo interesujący temat z obszaru robotyki. Odpowiednia fuzja informacji może pozwolić na skuteczniejsze podejmowanie decyzji w systemach bazujących na kilku, różnych źródłach. W celu głębszego zrozumienia tego zagadnienia, dokonujemy przeglądu istniejących starszych oraz nowszych metod integracji, a także ich obecnych zastosowań i potencjalnych problemów z nimi związanymi.

I. WSTĘP

Systemy uczenia maszynowego w robotyce mają do dyspozycji dane o różnych modalnościach. Często spotykanymi modalnościami są obraz, dźwięk i numeryczne szeregi czasowe z sensorów (np. pozycji, rotacja, siła, lidar).

Efektywne systemy uczenia maszynowego powinny maksymalnie wykorzystywać informacje w dostępnych modalnościach, rozpoznając zależności pomiędzy różnymi modalnościami.

Dodatkowo, jeśli różne modalności współdzielą informacje, to zaszumienie lub wyłączenie jednej lub kilku powiązanych modalności nie powinno bardzo negatywnie wpływać na jakość predykcji.

II. PRZEGLĄD ZAGADNIEŃ ORAZ WYNIKI

A. Luo, Chang [1]

W przeglądowym artykule, Luo i Chang wymieniają i porównują klasyczne podejścia do problemu fuzji danych. Jedną z nich to filtr Kalmana i jego rozszerzona wersja. Jako zalety filtru Kalmana autorzy wymieniają wydajność i prostotę implementacji. Minusami tych metod są ograniczające założenia.

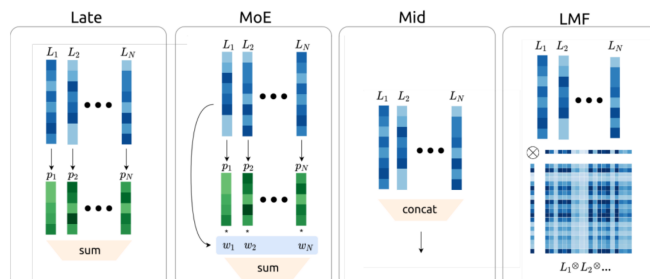
Autorzy wymieniają też metody oparte na kowariancji (covariance intersection / covariance union), które radzą sobie z danymi z różnych sensorów lepiej niż filtr Kalmana, oraz są odporne na niespójności pomiędzy modalnościami.

Inną istotną metodę fuzji, którą przedstawiają autorzy jest po prostu konkatenacja wektorów różnych modalności i użycie SVM np. do zadania klasyfikacji.

B. Bednarek, Kicki, Walas [2]

Autorzy kolejnego artykułu przyjrzeni się czterem metodom fuzji danych (widocznym na obrazku 1), które mogą być użyte jako część neuronowych modeli uczenia maszynowego.

Metoda Late jest najprostszą metodą fuzji danych. Łączy ona modalności przez sumę wektorów predykcji. Metoda MoE integruje dane z wielu modalności przez sumę ważoną wektorów predykcji dla każdej modalności. Wagi modalności w MoE zależą od reprezentacji ukrytej modalności. Metoda Mid konkatenuje reprezentacje ukryte. Połączone wektory mogą



Rysunek 1. Wizualizacja testowanych metod fuzji

być przetwarzane dowolnym modelem uczenia maszynowego. Warto zauważyć, że metoda Mid została opisana przez Luo i Chang (konkatenacja wektorów modalności i klasyfikacja przez SVM), ale nie została nazwana. Metoda LMF zamiast konkatenować reprezentacje ukryte, oblicza ich tensorowy produkt zewnętrzny. Uzyskany tensor można wykorzystać w dowolnym modelu uczenia maszynowego.

Autorzy porównali metody przy pomocy trzech zbiorów danych: BiGS (BioTac Grasp Stability Dataset), HaTT (Penn Haptic Texture Toolkit) i PHAC-2 (Penn Haptic Adjective Corpus 2). Zbiór BiGS został wykorzystany w zadaniu predykcji czy ręka robota upuści dany obiekt. Zbiór zawiera dane takie jak pozycja, obrót i siły ręki robota oraz dane z czujników dotyku. Zbiór HaTT został użyty w zadaniu rozpoznawania tekstur z szeregów czasowych z danymi haptycznymi. W zbiorze HaTT każda tekstura ma przypisany jeden obrazek, którego jednak autorzy nie użyli, ponieważ pojedynczy obrazek nie jest wystarczający do uczenia testowanych metod. Autorzy wykorzystali zbiór PHAC-2 do zbadania jakości wieloetykietowej klasyfikacji własności tekstur. Zbiór składa się ze zdjęć klasyfikowanych obiektów z różnych kątów oraz szeregów czasowych z danymi haptycznymi. Jako miarę jakości klasyfikacji wieloetykietowej dla zbioru PHAC-2, autorzy zastosowali AUC-ROC. W pozostałych zbiorach mierzono accuracy.

Rysunek 2. Jakość klasyfikacji przy zaszumieniu i zerowaniu danych z najważniejszej modalności

	Late	MoE	Mid	LMF
BiGS [%]	85.77	88.76	86.27	89.01
HaTT [%]	54.85	57.05	53.95	53.25
PHAC-2 [-]	0.23	0.24	0.21	0.61

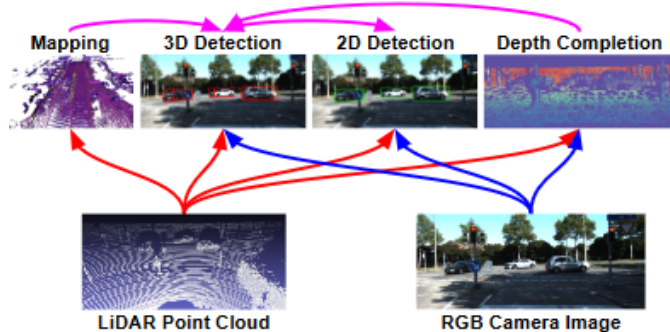
W eksperymentach nie zaobserwowano znaczących różnic pomiędzy jakością klasyfikacji architektur fuzji danych. Autorzy powtórzyli też eksperymenty dla klasyfikatorów uczonych na zaszumionych i częściowo wyzerowanych danych w najważniejszej modalności. Wybrane wyniki eksperymentu z zaszumianiem przedstawiono w tabeli 5. Jakość klasyfikacji była podobna dla różnych metod fuzji w zbiorach BiGS i HaTT. Natomiast w zbiorze PHAC-2, metoda LMF uzyskała znacząco lepszy wynik od innych metod.

C. Liang et al. [3]

Autorzy artykułu "Multi-Task Multi-Sensor Fusion for 3D Object Detection" przedstawili wykorzystanie fuzji sensorów w ciekawej i dosyć rozbudowanej architekturze end-to-end.

W celu rozwiązania problemu detekcji obiektów 3D wykorzystano dwa sensory: obraz z kamery RGB oraz chmurę punktów z pomiarów LiDAR. By skutecznie przeprowadzić trening tak zaawansowanej architektury, autorzy wspomogli naukę modelu, wykorzystując zadania pomocnicze: detekcję obiektów w 3d, estymację podłoża oraz estymację głębokości.

Rysunek 3. Diagram połączeń podzadań i danych z sensorów



Oprócz satysfakcjonujących wyników końcowych, ich eksperymenty doprowadziły do wniosku, iż zadania te są komplementarne, a ich wspólne wykorzystanie wspomaga sieć w tworzeniu lepszych reprezentacji w wyniku fuzji informacji.

Rysunek 4. Poprawa wyników w zależności od konfiguracji modelu

Model	Multi-Sensor		Multi-Task			2D AP (%)			3D AP (%)		
	pt	roi	map	dep	depf	easy	mod.	hard	easy	mod.	hard
LiDAR only						93.44	87.55	84.32	81.50	69.25	63.55
+image	✓					+2.95	+1.97	+2.76	+4.62	+5.21	+3.35
+map		✓				+3.06	+2.20	+3.33	+5.24	+7.14	+4.56
+refine	✓	✓	✓			+3.94	+2.71	+4.66	+6.43	+8.62	+12.03
+depth	✓	✓	✓	✓		+4.69	+2.65	+4.64	+6.34	+8.64	+12.06
full model	✓	✓	✓	✓	✓	+4.61	+2.67	+4.68	+6.40	+8.61	+12.02

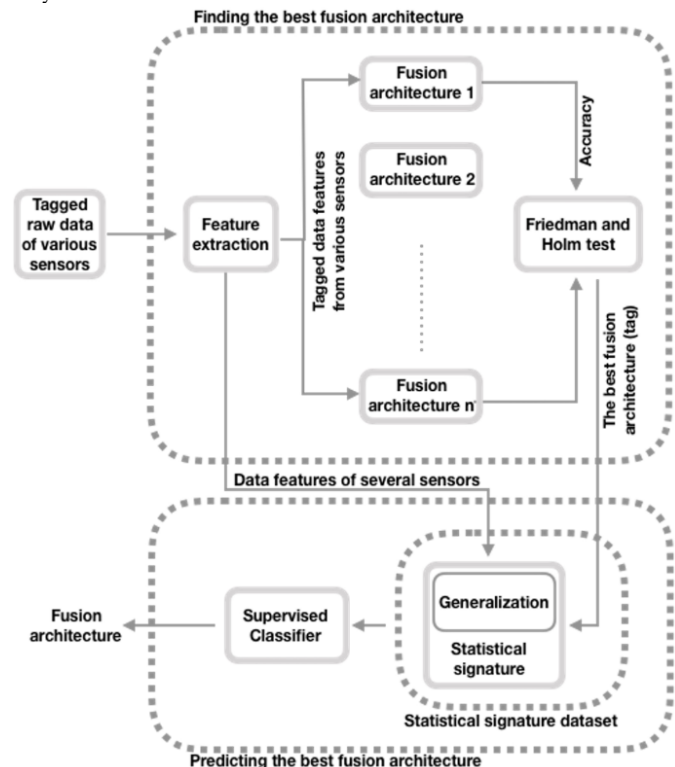
Warto zauważyć, iż mimo porównania wyników rozwiązania autorów do wyników state-of-the-art z momentu publikacji artykułu wykazującego przewagę zaproponowanej architektury, już rok później ich wyniki na benchmarkowym zbiorze danych KITTI zajmują miejsce 167 na problemie detekcji samochodów w 3d.

D. Molino-Minero-Re et al. [4]

Molino-Minero-Re wraz ze współpracownikami zauważyli nowe zagadnienie dotyczące fuzji danych z sensorów. Problem ten spotkał się bowiem z wysokim zainteresowaniem, w wyniku czego powstało wiele różnych podejść lub architektur znacząco różniących się od siebie. W efekcie nietrywialne było określenie, które ze znanych podejść jest najbardziej efektywne dla określonego zbioru sensorów i środowiska.

Autorzy zaproponowali zatem metodę, której celem jest dokonanie predykcji najlepszego z wybranych podejść do fuzji danych z sensorów, co powinno pozwolić na odejście od metody "prób i błędów".

Rysunek 5. Schemat metody identyfikującej optymalną architekturę fuzji danych z sensorów



Zaprezentowana metoda obejmuje między innymi budowę meta-zbioru danych, którego celem jest ekstrakcja cech statystycznych z oryginalnego zbioru danych. Takie działanie pozwala na stworzenie spójnego zbioru danych umożliwiającego trening algorytmu uczenia maszynowego. Wykorzystano w tym celu dwie metody: algorytm Sequential Forward Floating Selection oraz transformację T.

W wyniku przeprowadzonych eksperymentów wykazano, że na tak przetworzonych danych predykcja wykonana za pomocą algorytmu Losowych Lasów pozwala uzyskać satysfakcjonujące wyniki na poziomie 93%.

III. WNIOSKI

Integracja danych z wielu sensorów to bardzo obiecujący obszar, pozwalający na skuteczne wykorzystanie danych z

wielu źródeł w sposób efektywny, nie tracąc, a nawet zyskując na jakości wyników.

Zagadnienie to nie jest zupełnie nowe oraz istnieje już wiele interesujących metod bazujących między innymi na filtrze Kalmana, oraz kowariancji; dopiero teraz jednak, w dobie szybkiego rozwoju sztucznej inteligencji oraz uczenia głębokiego, nowe metody pozwalają ukazać prawdziwy potencjał skutecznej fuzji danych. Zaawansowane architektury osiągają coraz lepsze wyniki, swój potencjał zawdzięczając między innymi efektywnym wykorzystaniu dostępnych danych.

Podejrzewamy, że w przyszłości, badania na temat integracji danych z wielu sensorów jeszcze niejednokrotnie nas zaskoczą, biorąc pod uwagę, iż jest to jeden z kluczowych aspektów rozwoju autonomicznych pojazdów.

BIBLIOGRAFIA

- [1] Ren C Luo i Chih-Chia Chang. “Multisensor fusion and integration: A review on approaches and its applications in mechatronics”. W: *IEEE Transactions on Industrial Informatics* 8.1 (2011), s. 49–60.
- [2] Michał Bednarek, Piotr Kicki i Krzysztof Walas. “On Robustness of Multi-Modal Fusion—Robotics Perspective”. W: *Electronics* 9.7 (2020), s. 1152.
- [3] Ming Liang i in. “Multi-task multi-sensor fusion for 3d object detection”. W: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, s. 7345–7353.
- [4] Erik Molino-Minero-Re i in. “Improved Accuracy in Predicting the Best Sensor Fusion Architecture for Multiple Domains”. W: *Sensors* 21.21 (2021), s. 7007.