

교통사고 심각도 분류분석을 위한 데이터 융합

이성호* · 손소영*

*연세대학교 기전공학부

Abstract

데이터 융합(Data Fusion)은 하나의 센서에 의해 성취될 수 있는 것 이상의 추론과 개선된 정확성을 얻기 위해 여러 개의 센서로부터 감지된 데이터를 조합하는 기법, 이론, 시스템을 통틀어 말한다. 데이터융합의 응용분야로는 자동타겟인식, 전장감시, 전략적 경보방어시스템등의 군사분야 및 복잡한 기계의 모니터, 의학진단, CBM(condition-based maintenance)등의 비군사분야를 포함한다. 데이터융합은 데이터수준에 따라 특징레벨융합(feature-level fusion) 또는, 결정레벨융합(decision-level fusion)으로 나눌 수 있고, 관련 기법으로특징레벨융합에서는 시그널 프로세싱 기법(셋트럴분석, 시간-빈도분석, 비선형동적기법, 상위하위순위분석), 특징분류기법(뉴럴네트워크, 패턴인식, 퍼지로지, 클러스터링 알고리즘, 템플릿방법)등이 사용되고, 결정레벨융합에서는 가중결정방법, 전형적추론방법, 베이지안 추론, 덤스터-쉐퍼방법 등이 사용되고 있다.

본 논문에서는 교통사고 심각도 분류분석에 사용된 신경망과 Decision-Tree기법을 각각의 센서로 고려하고 그 기법들에 의해 실시간으로 입력되는 각 경우별 물적피해 또는, 신체상해 분류결과의 정확도와 분별력을 덤스터-쉐퍼와 베이지안 방법을 이용하여 향상시켜 보고자 한다. 두개의 센서(신경망, Decision-Tree)로 결정될 수 있는 기본제안을 각각 교통사고의 결과인 신체상해와 물적피해라 하였고, 기법(신경망, Decision-Tree)별 분류결과에 대한 믿음의 함수를 덤스터-쉐퍼방법을 이용하여 융합하였다. 또, 덤스터-쉐퍼방법의 결과를 베이지안방법의 사전확률값으로 사후확률(융합된 신뢰도)을 구하여 분류정확성 및 분별력 향상을 도모하였다. 이를 위하여 1996년 실제 교통사고 표본자료를 Training, Validation, Test 자료로 각각 40%, 30%, 30% 할당한 결과 3469개 Test 자료에서 덤스터-쉐퍼방법이 신경망보다 평균 3.65%, Decision-Tree보다 평균 8.88%의 분별력 향상을 나타냈다. 그리고, 1000개의 분류데이터를 제외한 2469개의 Test자료에서 베이지안방법이 신경망보다 평균 15.21%, Decision-Tree보다 평균 20.85%의 분별력 개선결과를 나타내 베이지안방법이 분별력면에서는 더 우수한 결과를 입증하였다. 그러나, 신경망의 분류정확도는 70.86%, Decision Tree는 72.3%였는데 반해, 덤스터-쉐퍼방법에 의하여 융합한 결과는 72.79%로 나타났다. 3469개의 Test자료중 1000개를 사용하여 조건부 확률을 구하고 나머지 2469개의 자료를 가지고 베이지안 방법에 의한 사후확률값을 이용하여 구한 분류정확도는 72.17% 였다. 2469개의 자료로 신경망(69.58%)과 Decision - Tree (72.22%), 덤스터-쉐퍼방법 (72.13%)에 의한 분류정확도를 구하여 비교해본 결과 전반적으로 큰차이는 없는 것으로 나타났다. 분류정확도를 향상시키기 위해 덤스터-쉐퍼방법으로 오분류수를 살펴본 결과 총3469개의 Test 자료 중 944개로 나타났다. 오분류가 발생한 형태를 살펴보니 실제 신체상해사고가 발생하였을 때 신경망, Decision-Tree, 덤스터-쉐퍼방법이 물적피해라고 분류한 비율이 총944개의 자료중 약 50%로 나타났고, 실제 물적피해사고가 발생하였을 때 세 기법이 모두 신체상해라 분류한 비율이 약30%로 나타났다. 이러한 오분류자료의 형태를 분석, 연구하여 분류정확도를 향상시키는 데이터융합 방법의 개발이 필요하다.