신경망과데이터베이스 관리시스템을 이용한 실시간 교통상황 예보

진현수
백석대학교 정보통신학부
e-mail:jhs1020@bu.ac.kr

Forecasting of Real Time Traffic Situation using Neural Network and Sensor Database Management System

Hyun Soo Jin
BaekSeok University

요 약

본 논문에서는 교통사고를 예방하고 교통사고 구간 대기시간을 줄이기 위해서 신경망을 이용한 예측방법을 제안한다. 뿐만 아니라, 교통사고 예측에 있어서 신경망에 정규화하지 않은 데이터를 사용하는 방법을 제시한다. 이 방법은 신경망 훈련 시 데이타의 최대 값을 추정할 필요가 없어 정규화 없이 신경망을 사용 가능하며, 신뢰성 예측 결과도 추정된 최대 값과 실제 획득될 최대 값과의 차이(추정 오차)만큼 정확해질 수 있다. 또한 비정규화 된 데이터를 사용하는 방법이 데이터의 최대 값을 알고 있다고 가정한 상태의 정규화된 방법보다 예측 정확성이 좋음을 보였다. 모의실험결과 제안된 신경망 예측시스템이 신경망을 고려하지 않은 기존방법보다 교통사고 구간 대기시간을 줄일 수 있음을 입증했다. 이와 같이 검증된 예측능력을 바탕으로 사용자에게 교통상황을 실시간으로 서비스하기 위하여 센서 데이터베이스를 이용한 실시간 교통정보 예보 시스템을 제안한다.

Abstract

This paper proposes a prediction method to prevent traffic accident and reduce to vehicle waiting time using neural network. Computer simulation results proved reducing average vehicle waiting time which proposed coordinating green time better than electro-sensitive traffic light system dose not consider coordinating green time. Moreover, we present neural network approach for traffic accident prediction with unnormalized (actual or original collected) data. This approach is not consider the maximum value of data and possible use the network without normalizing but the predictive accuracy is better. Also, the unnormalized method shows better predictive accuracy than the normalized method given by maximum value. Therefore, we can make the best use of this model in software reliability prediction using unnormalized data. Computer simulation results proved reducing traffic accident waiting time which proposed neural network better than conventional system dosen't consider neural network.

1. 서론

기존의 교통신호시스템은 대형차량의 구성비 가 적고, 과포화 상태가 적었던 1960 년도에는 년별 월별 일별로 교차로를 통행하는 평균차량 data를 미리 입력 시켜서 교통 신호주기를 생성 시키는 Time of Day(T.O.D.)신호등 방식도 큰문제가 없었으므로, Webster 방식이나, Box Jen kins 혹은 Kalman Filter을 이용한 교통예측 방법과, Box-Jenkins Process에 Moving Average 를 적용한 예측 모형을 이용하였다[1,2,3].

최근 들어 고장구간 모델의 모수 추정과 모델 자체 학습으로 미래의 결과를 예측하는데 신경망 이 적용되고 있다 .이러한 고장구간을 예측하려면, 미리 입력된 예상 교통 사고구간 데이타에 적합하도록 학습을 통해 자율적으로 망의 내부 모델을 개발하고 미래의 교통사고 사고구간 발생 과정에 대한 예측을 할 수 있다. [8],[9],[10],[14].

2장에서는 교통사고구간 자동판단에 관해서 살펴본다. 3장에서는 교통사고를 예측하는 신경망방식을 설명한다.

4장에서는 신경망을 이용한 교통사고분석에 관해서 알아본다. 5장에서는 교통사고가 발생했을 때에 기존의 방법으로 예측 및 사고처리방법과 퍼지 신경

망을 이용한 교통실시간 예보 시스템의 교통사고 발생시, 사고구간 판단 및 복구 시간을 비교 분석한다.

2. 교통 사고구간 자동판단

본 논문에서는 신경망 및 퍼지 전문가시스템을 이용하여 폭설구간, 안개구간, 결빙구간, 교통사고 구간, 등을 판단하는 알고리즘 및 기존 방법과 비교해서 교통사고발생시 사고처리에 소요되는 대기시간 결과를 제시한다.

교통사고 신뢰성을 추정하거나 예측하는데 사용된다. 일반적으로 수집되는 고장 데이타 형태는 고장수 (일정한 시간 간격동안 발생한 고장 수 기록)나 고장 시간 (각 고장이 발생한 시점을 기록하는 고장시간 데이타) 이며, 고장시간 데이타 형태가 아닌 고장시간 데이타를 그룹화한 형태인 그룹 데이타 (Grouped data) 형태만 활용할 수 있는 경우도 있다. 이 데이타 형태는 고장 수 데이타와 동일하나 다른 점은 가변적인 시험 간격 (Variable time interval of Testing) 형태를 취한다.

고정된 시간 간격에서 수집된 고장 수를 이용할 경우, 현재 시점의 신뢰성을 추정하거나 고정된 미래 시점의 신뢰성을 예측하는데는 문제가 되지 않는다. 그러나 그룹 데이타의 경우, 가변적인 시간 간격을 가지므로 고정된 시간 간격에 적합한 모델로 가변적인 미래 시점에서의 신뢰성을 예측하는데는 문제가 발생될 수 있다. 기존의 통계적 SRGMs 들은 가변 시간 간격을 가진 그룹 데이타에 대한 많은 연구가 이루어지지 못하였으며, 특히 신경망 모델에 대한 연구는 없는 실정이다.

따라서, 본 논문에서는 가변적인 시간 간격을 가진 그룹 데이타에 적합한 신경망 모델을 개발하는 연구를 수행하고자 한다.

교통사고 및 위험구간 데이타 쌍들 $(t_i, C_i), i=1,2,\cdots,n$ 이 소프트웨어 시험 결과 관찰되었다고 가정하자. 여기서 C_i 는 i 번째 시험 시간인 t_i 까지 발견된 누적 고장 수를 의미하며, 등 간격을 유지하는 단위 시험시간 $\Delta t_i = t_i - t_{I-1}$로, i번째 단위 시험시간에서 발견된 고장 수 $\Delta C_i = m_i = C_i - C_{i-1}$로 정의된다. 이들 데이타 쌍들을 고장 수 데이타라 칭한다.

소프트웨어 신뢰성 분야에서, 미래에 발생되는 고장 수를 예측하는데 가장 널리 알려진 망으로는 FFNs와 PRNs (Partial Recurrent Networks)인 Jordan과 Elman 망이 있다. Karunanithi et al. [8],[9]는 신경망을 활용하여 소프트웨어 신뢰성 예측을 모델링하였으며, 소프트웨어 신뢰성 예측에 캐스케이드-상관 학습 알고리즘 (Cascade-correlation learning algorithm : CC)을 이용하여 FFNs과 PRNs이 적용될 수 있음을 보였다. 신경망의 예측력을 평가하기 위해 2가지 훈련 제도 (Training regimes)를 고려하였다.

일반 훈련제도(Generalization training regime)
: 입력 (Input) t_i는 목표 (Target) C_i에 관련
예측 훈련제도(Prediction training regime)
: 입력 t_{i-1}는 목표 C_i에 관련

또한, 고장 수 데이타에 대해 잘 알려진 통계적 SRGMs들과 비교하였다. 실험에 사용된 14개의 데이타를 살펴보면, 10개 (Data1, Data3-10 와 Data14)는 고장 수 데이타이며, 나머지 4개 (Data2와 Data11-13)는 고장시간 데이타이다. 고장 시간 데이타에 대해 고장 수 데이타 신경망 모델을 잘못 적용한 점을 보완하기 위해 Park et al. [16]은 고장 시간 데이타에 대해 유사한 연구를 수행하였다. 10개의 고장 수 데이타 중 Data4와 Data5는 데이타 수집 시간 간격이 일정하지 않은 그룹 데이타의 경우이고 나머지들은 데이타 수집 시간 간격이 일정한 경우이다.

신경망을 함수 근사, 회귀분석 또는 시계열 분야에 이용하는 경우, 주어지는 시간 간격은 일정한 데이타를 활용하는 것이 일반적이며, 또한 소프트웨어를 시험하면서 수집되는 고장 수 데이타는 일반적으로 시간 간격이 일정한 형태를 취한다. 그러나 <표 1>과 같이 데이타 수집 시간 간격이 가변적인 그룹 데이타의 경우도 발생한다. 가변 시간간격을 가진 그룹 데이타를 기존의 고정된 시간간격에 적합한 신경망 모델을 이용하여 미래의 가변 시간 간격 시점에서의 고장 수를 예측할 경우 문제가 발생한다. 예를 들면, Karunanithi et al. [8],[9]이 제안한 예측 훈련제도의 경우 입력은 t_{i-1}인데 목표는 고정된 시간간격인 t_i 시점에서의 누적 고장 수인 C_i이다. 따라서, 그룹 데이타의 가변적 시간간격인 t_i 시점까지의 누적 발견 고장 수를 예측하려면 가변 시간간격 $\Delta t_i = t_i - t_{i-1}$ 정보를 알아야 한다. 그러나 기존 모델링에서는 Δt_i 정보가 전혀 들어있지 않음으로 문제가 발생하며, 이 모델은 Δt_i가 일정할 때만 적용 가능하다. 이때는 t_i 들을 i로 대체하여도 동일한 결과를 얻을 수 있다. 즉, Δt_i가 가변 시간 간격일 경우, 기존 신경망 모델

로는 Δt_i 시간 후에 발견되는 고장 수는 예측이 불가하다. 이 문제를 해결하기 위해서는 현재 시점인 t_{i-1}에서 얼마 동안의 시간에 발생하는 고장 수를 예측하는가 하는 가변 시간간격 (Δt_i)이 주어져야 한다. 이러한 그룹 데이타에 적합한 신경망 연구는 수행되지 않았다. 따라서, 가변적인 시간 간격 데이타에 적합한 신경망 모델을 얻기 위해 수집된 데이타를 이용하여 Karunanithi et al. [8],[9]이 제안한 훈련 제도를 변경시켜 적절한 신경망 모델을 제시하고자 한다.

<표 1> 그룹 데이타 특성 및 고장 데이타

데이타	참고문헌	LOC	고장수	Data Size	Application (Time interval)
Data 4	[15]	1,317,000	328	17	PL/I database application software (CPU execution time in Hours)
Data 5	[15]	35,000	279	10	Hardware control program (Time of Observation in Months)
System T1	[14]	21,700	136	21	Real time command control (CPU execution time in Hours)
System T38	[14]	Not known	32	11	Not known (CPU execution time in Hours)

Data4				Data5			
t_i	Δt_i	ΔC_i	C_i	t_i	Δt_i	ΔC_i	C_i
2.45	2.45	15	15	1.0	1.0	10	10
4.9	2.45	29	44	1.5	0.5	64	74
6.86	1.96	22	66	2.0	0.5	18	92
7.84	0.98	37	103	3.0	1.0	43	135
9.52	1.68	2	105	4.5	1.5	44	179
12.89	3.37	5	110	6.0	1.5	13	192
17.1	4.21	36	146	8.0	2.0	34	226
20.47	3.37	29	175	11.0	3.0	28	254
21.43	0.96	4	179	12.0	1.0	15	269
23.35	1.92	27	206	13.0	1.0	10	279
26.23	2.88	17	223				
27.67	1.44	32	255				
30.93	3.26	21	276				
34.77	3.84	22	298				
38.61	3.84	6	304				
40.91	2.3	7	311				
42.67	1.76	17	328				

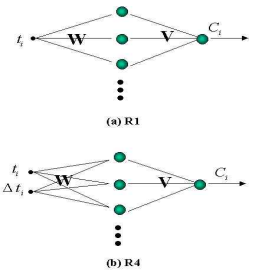
3. 교통사고 예측 신경망 모델

소프트웨어 시험 또는 운영 중에 I 시점까지 획득된 가변 시간간격 고장 수 데이타 (t_i, C_i), $i=1,2,\cdots I$를 활용하여 미래의 가변 시간 간격 시점 t_{I+r}에서의 고장 수 C_{I+r}를 신경망을 이용하여 예측하는 경우, 최적의 신경망을 어떻게 선택할 것인가가 문제로 제기된다. 본 장에서는 이러한 문제점을 해결하는 방법을 제시한다.

<표 2>에서 R1과 R2는 각각 Karunanithi et al. [8],[9]가 제시한 일반 훈련제도와 예측 훈련제도에 해당되며, R3-R13은 새로 제안된 훈련제도이다. 새로 제안된 훈련제도는 다음과 같이 표현된다.

· R3 : 입력 t_i와 t_{i-1}는 목표 C_i에 관련
· R4 : 입력 t_i와 Δt_i는 목표 C_i에 관련
⋮
·R13 : 입력 t_{i-1}, Δt_i와 ΔC_i는 목표 C_i에 관련

즉, Karunanithi et al.이 제안한 훈련제도 1 (R1)과 본 논문에서 제안하고 있는 훈련제도 4 (R4)에 대한 신경망의 구조는 (그림 1)과 같다.



(그림 1) 훈련제도 1 (R1)과 4 (R4)의 신경망 구조

4. 결론

본 논문에서는 교통사고구간 고장 데이터를 정규화 하지 않은 원 데이터를 신경망에 적용하여 교통사고구간을 예측하는 분야에 관해 연구하였다. 망 구조로는 FFN을 선택하였으며, 적절한 은닉 뉴런 수를 결정하기 위해 관찰된 데이터를 훈련과 검증 데이터로 분류하고 관찰되지 않은 다음 단계 데이터를 시험 데이터로 하여 예측 오차를 측정하는 방법으로 최적시점 종료기법을 적용하였다.