

고속도로 통행료수납자료를 이용한 통행시간 군집현상에 관한 연구

A Study on Travel time Platoon Formation using the data from Toll Collection System

박원식*, 최진우, 양영규
Won-sik Park*, Jin-woo Choi, Young-kyu Yang
경원대학교 전자계산학과
pws1981@paran.com, cjw49@paran.com, ykyang@kyungwon.ac.kr

요약

본 연구에서는 고속도로와 같은 연속류에서의 통행시간 군집화 현상을 연구하여 신뢰성 있는 단위 시간 당 구간 대표 통행시간을 구하는 전처리 방법을 제시하는데 목적이 있다. 현재까지는 단위시간 당 구간의 통행시간 대푯값으로 하나의 평균값(Mean), 최빈값(Mode), 또는 중앙값(Median)이 사용되었다. 이의 문제점은 운전자의 주행 습관(빠른 주행, 느린 주행), 휴게소 이용, 도로 정체 등 다양한 요인으로 인하여 차량별 구간 주행 속도 간 편차가 많아 현재 사용하는 1개의 대푯값으로는 전체 차량의 운행 특징을 정확히 표현하기가 곤란하다는 점이다. 이를 개선하기 위하여 본 연구에서는 군집 방법을 이용하여 차량군을 복수의 비슷한 군집으로 나누고 나누어진 그룹별로 통행시간 대푯값을 선정하는 방법을 제시하고 실험하여 이 방법이 효과적임을 증명하였다.

1. 서론

최근 교통정보의 가치와 수요가 과거 어느 때 보다도 급증하고 있으며 이용자 입장에서 통행시간 정보가 필수적인 정보로 자리 잡고 있다. 또한 점점 복잡해져가는 고속도로망 확대로 동일한 출발/도착지를 갖는 통행에 있어 선택 가능한 경로의 수가 많아지고 있다.

이러한 구간통행시간 정보는 운전자 자신이 가고자 하는 목적지까지의 경로를 선택하는데 직접적인 결정 요소 중 하나이다.

최근 ITS 기술 발전에 따라 고속도로 요금 징수를 위하여 개발된 TCS(Toll Collection System)자료가 통행시간 분석에 사용되고 있다. TCS 통행시간 자료는 고속도로를 통행한 모든 차량이 경험한 통행시간의 전수 자료로서 정확도나 활용도가 매우 높을 뿐만 아니라, 별도의 설비투자

가 필요치 않다는 장점을 가지고 있다.

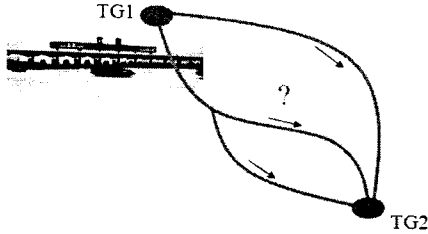
고속도로를 운행하는 운전자는 도로상황에 큰 변화가 있지 않아도 차량들의 일부가 군집하여 이동하는 현상을 경험한다. 차량들이 군집으로 이동하면 각 구간의 통행시간 또한 군집의 분포를 나타낸다.

운전자가 통행을 시작할 시점에서의 교통상황 정보는 운전 중에 달라지는 상황을 반영할 수 없어 운전자가 경험하는 실제 통행시간과 큰 차이를 보일 수 있다. 이러한 문제점을 해결하기 위해 통행시간을 군집으로 설정해 정상적인 소통흐름과 휴게소 이용, 여러 교통상황을 반영한 통행시간의 정보를 운전자에게 제공하는 것이 필요하다.

2. TCS 통행시간의 문제점 및 개선방안

<그림 1>는 TCS 데이터의 다중경로 환

경을 보여준다. TCS 자료가 통행시간정보로서 활용되기 위해서는 다음 2가지의 특성이 반드시 고려되어야 한다.



<그림 1> TCS 다중경로환경

첫째, 다중경로환경에서 TCS 통행시간은 경로통행시간이 아니다. TCS 통행시간은 고속도로 이용차량의 톨게이트 진입시각 및 진출시각의 차를 통해 계산된다. 그러나 진출입 톨게이트 간에 2개 이상의 경로가 이용가능 할 때에는 해당 TCS 통행시간이 어느 경로에 대한 것인지 알 수가 없다. 둘째, TCS 통행시간에는 고속도로 본선을 주행한 시간외에 갓길 정차시간, 휴게소 이용시간, 여러 도로교통 상황 등의 포함되어 있다. 이는 톨게이트간 거리가 멀어질수록 심해지는 경향이 있다. 이와 같은 두 가지 특성은 TCS 자료를 경로 통행시간정보로 활용하는 것을 어렵게 만드는 요인이 되어왔다.

이러한 TCS 자료를 이용한 경로 통행시간 예측 등 분석을 위한 대푯값 선정에는 각 구간별로 하나의 평균값(Mean), 최빈값(Mode), 중앙값(Median)이 사용되고 있다. 이의 문제점은 운전자의 운전습관(고속주행, 저속 주행), 휴게소 차량, 도로 정체 등 다양한 요인으로 인하여 차량별 구간 주행 속도 간 편차가 많다는 점이다. 따라서 현재 사용하는 1개의 대푯값으로는 전체 차량의 운행 특성을 정확히 대표하기가 곤란하다. 해결 방법으로는 각 구간 별 통행시간을 1개의 대푯값으로 표현하기 보다는 운행 특성 및 통행시간이 비슷한 차량들을 군집으로 그룹핑 하여 복수의 대푯값을 군집 별로 설정하여 활용하는 것이 효과적이다.

2.1 군집분석

군집분석(cluster analysis)은 모집단 또는 범주에 대한 사전정보가 없는 경우에 관측치(observation)들 사이의 유사성-거리 혹은 상관관계를 이용하여 전체를 몇 개의 군집(Cluster)으로 나누는 통계학적 분석기법이다. 결과적으로 그룹 내에서 동일한 특성을 나타내고, 그룹 간에는 서로 상이한 특성을 나타내도록 모집단을 여러 그룹으로 나누는 방법이라 할 수 있다.

군집분석의 방법은 각 관측치 사이의 유사성의 척도로써 무엇을 사용하느냐에 따라 여러 가지로 분류될 수 있다. 유사성의 척도로 사용되는 것에는 각 측정치 사이의 상관계수와 거리(distance)가 있다. 유사성의 척도로 거리에 의한 것이 많이 사용되는데, 두 객체간의 속성이 유사할수록 유사성거리는 가깝게, 두 객체간의 속성이 상이할수록 유사성 거리는 멀게 측정되어야 한다. 거리의 측정 방법으로서 본 연구에서는 유클리디안 거리를 사용하였다.

유사성의 척도에 기초해 그룹 내의 변량에 대한 그룹간의 변량을 최소화시키는 알고리즘에 따라 계층적(hierarchical)방법과 비계층적(nonhierarchical or disjoint)방법, 중복(overlapping)군집, 퍼지(fuzzy)군집방법이 있다. 계층적 방법은 유사성이 가까운 개체끼리 차례로 묶거나 멀리 떨어진 개체를 차례로 분리해 가는 군집방법으로서, 한번 병합된 개체는 다시 분리되지 않는 것이 특징이다. 계층적 방법에는 최단거리법(nearest neighbor method), 최장거리법(furthest neighbor method), 메디안법(median method), 중심법(centroid method), 군평균법(group average method), 워드법(Ward method)가 있다.

본 연구에서는 최단거리법과 중심법을 사용하였다.

최단거리법과 중심법을 비교해 보면 단일연결법에서는 군집간의 유사성 척도로 두 군집의 모든 객체 쌍의 거리 중 가장 가까운 거리를 사용한다. 단일연결법은 두 군집의 유사성을 가장 짧은 거리를 갖는 객체 쌍으로 평가하고자 하는 것이다. 이

를 최단거리법이라고도 한다. 중심법은 군집을 이루는 객체들의 중심이 되는 좌표를 그 군집의 중심(centroid)라 하는데, 중심연결법은 군집간의 유사성 척도로 두 군집의 중심간 거리를 사용한다.

대부분의 군집 방법은 특이값(outlier)에 굉장히 민감하므로 특이값 들은 사전에 면밀히 검토하여 제거하는 것이 바람직하다. 본 연구에서는 계층적 군집분석의 최단거리법과 중심법에 의한 군집분석 결과를 서로 비교하여 군집의 타당성을 검토하였다.

객체간의 유사성의 정도를 정량적으로 나타내기 위해서 척도가 필요하다. 가장 보편적으로 많이 사용되는 것이 거리(distance)인데 거리와 같이 클수록 유사성이 적어지는 척도는 보다 엄밀하게 비유사성 척도(dissimilarity measure)라 한다.

각 객체가 p개의 속성 또는 변수(variable)를 갖는다 하고 객체 i의 j번째 변수를 라 하면 객체 i의 p차원 공간에서의 좌표는 다음과 같은 열벡터로 표현된다.

$$d_{rs}^2 = \sum_{k=1}^p (x_{rk} - x_{sk})^2 \quad (1)$$

유클리디안 거리

군집분석 알고리즘은 다음과 같다.

- (1) 각 객체를 하나로 간주한다.
- (2) K=N
- (3) 현재의 군집 결과에 있는 모든 군집간의 쌍에 대하여 군집 간 유사성척도를 산출하며, 이 중 최소가 되는 군집 i와 군집 j를 묶어 하나의 군집으로 만든 후 군집결과를 수정한다.
- (4) k=1이면 Stop, 그렇지 않으면 단계 1을 반복한다.

3. 관련연구 고찰

이의은 외(2003)는 TCS 자료의 해당 수집주기의 대뿔값을 최빈값으로 설정하였다. 이는 여행시간 추정 및 예측 시 단위 시간 수집 자료군의 대표치 추출방법으로,

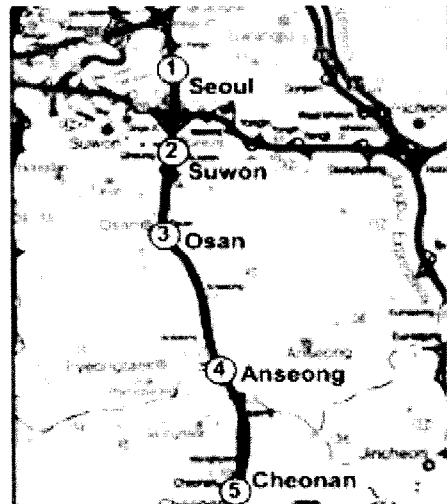
최빈값이 평균값보다 우수하다는 기존의 연구결과에서 참조하였다.(이승환외, 2000)

남궁성(1999)은 TCS 자료의 이상치제거 방법으로 누적상대도수비를 계산하여 상한치 95%값을 설정하였다. 대뿔값은 평균값으로 설정하였다.

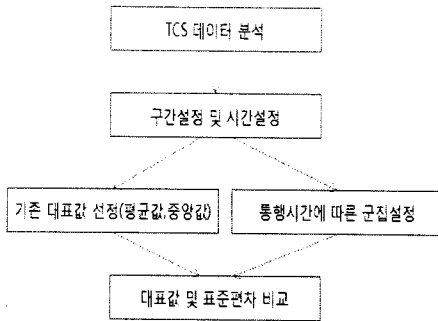
김재진 외(2006)는 구간검지시스템에서 수집되는 개별차량 통행시간 자료는 시스템상의 오류 및 정상적으로 주행한 차량이 아닌 차량들에 대한 이상치 들이 포함되어 있어 중위절대편차(MAD:Median Absolute Deviation)를 적용하였다.

4. 연구방법 및 자료

본 연구에서의 분석대상으로 서울-천안 구간을 선정하였다. 자료는 2007년 7월 1일 ~ 7월 2일간 08:00-10:10중 수집된 TCS 통행시간자료를 활용하였고 군집분석을 하기 위하여 SPSS 14.0을 사용하였다.



<그림 2> 경부선 서울-천안IC



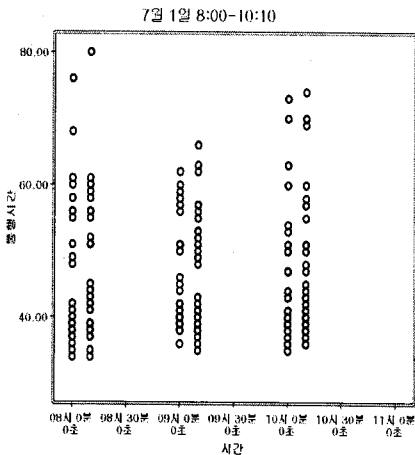
<그림 3> 분석

차종은 서울-천안구간 중 TCS의 1종에 해당하는 표준차종인 승용차 기준으로 데이터를 선택하였다.

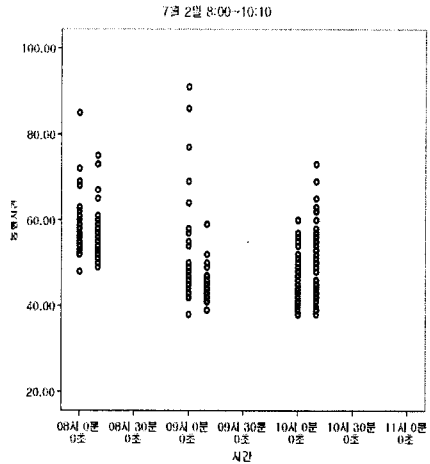
연구의 방법은 먼저 <그림 3>에서 보는 바와 같이 먼저 통행시간 산점도를 그려 단위 시간별 차량의 통행 특성 및 군집 형태를 분석한다. 이어 상기 2.1절에 기술한 알고리즘에 따라 군집분석을 행하고 기존 대푯값을 구하여 표준편차를 기준으로 효율성을 비교한다.

<그림 4>와 <그림 5>는 서울-천안 고속도로 개별 차량의 고속도로 통행시간 산점도를 보여준다. 각각의 개별 차량들의 통행시간(10분 단위)은 분포가 다양하다.

7월 1일, 7월 2일 08:00-10:00 통행시간 산점도를 보면 1개의 구간 대푯값을 설정하기에는 통행시간 분포가 다양함을 보여준다.



<그림 4> 7월 1일 08:00-10:10



<그림 5> 7월 2일 08:00-10:10

5. 분석결과 및 평가

먼저 서울-천안 구간 7월 1일 08:00-10:00 사이 서울을 출발한 차량들을 출발 시간 기준으로 10분 간격으로 모아 기존의 대푯값 산정방식의 평균값을 구한 결과는 <표 1>과 같다. <표 1>에서 8:00-8:10, 9:00-9:10, 10:00-10:10 시간대 별 구간 평균 통행시간은 각각 45.19, 44.96, 45.91분이고 표준편차는 10.70, 7.54, 9.54분으로 8:00-8:10 시간대가 가장 높은 표준편차를 보이고 있다.

시간 간격	운행횟수	구간 평균 통행시간(분)	표준편차(분)
8:00-8:10	31	45.19	10.70
9:00-9:10	33	44.96	7.54
10:00-10:10	36	45.91	9.54

<표 1> 7월 1일 기존방법(평균값)

한편 제안 알고리즘인 두 가지 군집방법(최단거리법, 중심법)으로 차량을 3개 군집으로 그룹화 하여 분석한 결과는 각각 <표 2>, <표 3>과 같다.

먼저 <표 2>에서 보면 차량의 운행 특성에 따라 차량의 차량이 3개 그룹으로 나뉘어 평균 통행 시간 및 표준편차가 계

산되었음을 보여준다. 8:00-8:10 시간대의 군집 2와 군집 3은 1개의 통행시간 값만 포함되어 있어 표준편차는 0이다. 군집 방법 중 최단거리법을 사용할 경우 군집 1에 대부분의 차량들이 포함되어지므로 군집1을 대푯값으로 간주할 수 있다. 최단거리법에 의한 각 시간대별 표준편차를 살펴보면 08:00-08:10, 9:00-9:10, 10:00-10:10 시간 간격별 군집 1의 표준편차는 각각 8.41, 2.29, 5.31분으로 <표 1>의 기존의 방법에 의한 표준편차보다 감소되었음을 알 수 있다.

시간	군집 구분	갯수	최소 값	최대 값	평균 (분)	표준 편차 (분)
8:00 8:10	1	29	34	61	43.3	8.41
	2	1	68	68	68	0
	3	1	76	76	76	0
9:00 9:10	1	23	36	46	40.3	2.29
	2	4	50	51	50.7	0.5
	3	6	56	62	58.6	2.16
10:00 10:10	1	31	35	54	42.7	5.37
	2	3	60	63	62	1.73
	3	2	70	73	71.5	2.12

<표 2> 7월 1일 군집법(최단거리법) 결과

중심법을 사용한 군집 분석의 결과는 최단거리법을 사용한 군집 결과보다 훨씬 개선된 결과를 보여준다.

<표 3>에서 보면 그룹1에 대부분의 차량이 배분되었던 단거리법에 비하여 세 개의 군집별로 비교적 균등한 차량 배분이 이루어졌음을 알 수 있다. 7월 1일 중심법 결과의 표준편차를 보면 8:00-8:10, 9:00-9:10, 10:00-10:10 시간대의 군집 1의 표준편차는 각각 1.92, 2.29, 3.40분으로 <표 1>의 시간대 별 표준편차와 <표 2>의 군집 1 표준편차와 현격한 차이

를 보이고 있다. 이는 중심법 군집 분석 결과에 따른 대푯값이 단일 평균값 및 최단거리 군집법보다 훨씬 신뢰도가 높은 것을 의미한다.

시간	군집 구분	갯수	최소 값	최대 값	평균 (분)	표준 편차 (분)
8:00 8:10	1	20	34	42	38.1	1.92
	2	9	48	61	54.8	4.64
	3	2	68	72	72	5.65
9:00 9:10	1	23	36	46	40.3	2.29
	2	4	50	51	50.7	0.5
	3	6	56	62	58.6	2.16
10:00 10:10	1	25	35	47	40.6	3.40
	2	6	50	54	51.5	1.64
	3	5	60	73	65.8	5.44

<표 3> 7월 1일 군집법(중심법) 결과

동일한 방법으로 7월 2일 자료를 분석해 보면 다음과 같다.

<표 4>에서 보면 8:00-8:10, 9:00-9:10, 10:10 시간대 구간 평균 통행시간은 각각 58.04, 49.63, 44.53분이고 표준편차는 각각 6.17, 11.67, 6.12분으로 9:00-9:10 시간대가 최대 편차를 보이고 있다.

시간 간격	운행갯수	구간 평균 통행시간(분)	표준편차 (분)
8:00 8:10	47	58.04	6.17
9:00 9:10	41	49.63	11.67
10:00 10:10	68	44.53	6.12

<표 4> 7월 2일 기존방법(평균값)

<표 5>에서 최단거리법을 적용한 군집 분석을 7월 1일 경우와 마찬가지로 각 시간대별로 동일하게 군집1에 대부분의 차량

들이 배분되었음을 보여준다.

7월 2일 자료를 최단거리법으로 군집한 결과를 보면 8:00-8:10, 9:00-9:10, 10:00-10:10 시간대 별로 그룹1의 평균 통행 시간은 각각 56.3, 46.8, 42.5분이고 표준편차는 3.23, 6.40, 3.74분으로 <표 4>의 기존 방법에 비하여 50%정도로 감소하여 상당히 개선되었음을 알 수 있다.

시간	군집 구분	갯수	최소값	최대값	평균 (분)	표준 편차 (분)
8:00 8:10	1	42	42	63	56.3	3.23
	2	4	68	72	69.2	1.89
	3	1	85	85	85	0
9:00 9:10	1	38	38	69	46.8	6.40
	2	1	77	77	77	0
	3	2	86	91	88.5	3.53
10:00 10:10	1	61	38	52	42.5	3.74
	2	4	54	57	55.5	1.29
	3	3	60	60	60	0

<표 5> 7월 2일 군집(최단거리법) 결과

중심법에 의한 군집분석 결과를 살펴보면 8:00-8:10, 9:00-9:10, 10:00-10:10 시간대 별로 군집1의 평균통행시간은 각각 56.3, 45.7, 41.8분이고 표준편차는 3.23, 4.44, 2.90분으로 <표 4>의 기존의 방법보다 훨씬 신뢰성이 높다. 최단거리를 이용한 군집 분석 결과와 비교하여 보면 8:00-8:10 시간대와 10:00-10:10 시간대에서는 같은 결과를 보여주나 9:00-9:10 시간대의 경우는 최단거리를 이용한 군집 분석 결과보다 훨씬 감소

된 편차 값을 보여준다. 이는 위에서도 언급한 바와 같이 그룹1에 대부분의 차량이 배분되었던 최단거리법에 비하여 제2그룹, 제3그룹에 최단거리법에 비하여 더 많은 차량 배분이 이루어졌기 때문인 것으로 보인다.

시간	군집 구분	갯수	최소값	최대값	평균 (분)	표준 편차 (분)
8:00 8:10	1	42	42	63	56.3	3.23
	2	4	68	72	69.2	1.89
	3	1	85	85	85	0
9:00 9:10	1	36	38	58	45.7	4.44
	2	3	64	77	70	6.55
	3	2	86	91	88.5	3.53
10:00 10:10	1	56	38	48	41.8	2.90
	2	5	49	52	50.8	1.30
	3	7	54	60	57.4	2.52

<표 6> 7월 2일 군집(중심법) 결과

상기 분석에서 보는 바와 같이 기존 대푯값의 표준편차와 복수의 군집 설정한 대푯값의 표준편차의 차이를 비교하면 복수의 군집으로 설정한 표준편차가 훨씬 적은 것을 알 수 있다.

이와 같은 분석에서 알 수 있듯이 운전자의 개별 형태를 감안한 평균통행시간을 제공하는 것이 보다 신뢰성 있고 효과적인 통행시간 제공 방법이라 할 수 있을 것이다.

6. 결론

TCS를 이용한 교통소통 정보 제공을 위하여 사용되고 있는 단일 대푯값의 문제점을 살펴보고 이의 해결방안으로 군집 방법을 통해 군집된 차량 그룹별 대푯값을 구하는 방법을 제안하고 실험하였다.

군집방법은 최단거리법과 중심법을 활용하였으며 서울-천안 간 구간에 10분 간격의 시간대 별 대푯값의 표준편차를 비교하여 분석 알고리즘의 효율성과 신뢰도를 분석하였다.

기존 방법과 제안 알고리즘의 표준편차의 범위를 보면 7월 1일-7월 2일 기존 대푯값의 범위는 6.12~11.67를 보여주고 있고 군집 설정한 표준편차의 범위는 1.20~6.40의 값을 나타낸다.

이는 결론적으로 제안된 알고리즘인 복수의 군집설정 방법이 기존의 1개의 대푯값 방법에 비하여 훨씬 효율적이고 신뢰성 높은 분석 방법임을 나타내준다.

향후 통행시간 정보의 질(Quality)측면에 정확성과 신뢰성을 위해 군집으로 설정, 운전자의 맞춤 통행시간 정보를 제공하기 위한 심도 있는 연구가 필요하다.

"본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구 결과로 수행되었음" (IITA-2008-C1090-0801-0040)

"본 논문은 한국도로공사 OASIS (Operations Analysis and Supportive Information System)의 자료지원을 받아 수행되었습니다.

참고문헌

[1] 김재진, 노정현, 박동주, "구간검지 체계를 이용한 On-Line 출발사각기준 링크 통행시간 추정(연속류를 중심으로)", 대한교통학회지, 1229-1366, 제24권2호 pp,157-168, 2006

[2] 남궁성, "고속도로 경로통행시간 산출을 위한 전진반복 전후방탐색법(PIFAB)

의 개발", 대한교통학회지, 1229-1366, 제23권5호, pp.147-155, 2005

[3] 이의은, 김정현, "시간처짐현상을 고려한 장거리구간 통행시간 예측 모형 개발", 대한교통학회지, 1229-1366, 제20권4호, pp.51-61, 2002

[4] 강정규, 남궁성, "고속도로 통행료수납자료를 이용한 통행시간 예측모형 개발", 대한교통학회지, 1229-1366, 제20권4호, pp.55-67, 2002

[5] 이승재, 백남철, 권희정, "단기조사 교통량을 이용한 AADT 추정연구", 대한교통학회지, 1229-1366, 제20권 6호, pp.59-68, 2002

[6] 김성현, 임강원, 이영인, "일반국도 통행시간 추정을 위한 동질구간 기반 지점검지기 배치에 관한 연구", 대한교통학회지, 1229-1366, 제24권 1호, pp, 73-84, 2006

[7] 남궁성, "TCS 자료를 이용한 고속도로 통행시간 예측", 대한교통학회지, 제34회 학술발표회, pp.489-494, 1999

[8] 이준, 정진혁, "고속도로 교통류의 차량군집현상에 관한 연구", 대한교통학회지, 제25권 제2호, pp.109-120, 2007