

GeoSensor 환경에서 공간 질의 정확도 향상을 위한 선-필터링을 이용한 후-부하제한 기법

(Pre-Filtering based Post-Load Shedding Method for Improving Spatial Queries Accuracy in GeoSensor Environment)

김 호* 백 성 하** 이 동 욱** 김 경 배*** 배 해 영****
(Ho Kim) (Sung Ha Baek) (Dong Wook Lee) (Gyoung Bae Kim) (Hae Young Bae)

요 약 u-GIS 환경에서 GeoSensor를 기반으로 하는 GeoSensor 환경은 다양한 센서들로부터 수집한 동적인 데이터와 기존 GIS인 정적인 지형지물 정보의 융합을 요구한다. 이 환경의 핵심인 GeoSensor는 넓은 지역에 산발적으로 분포하며, 다양한 크기의 데이터를 끊임없이 수집한다. 따라서 Data Stream Management System(DSMS)은 제한된 메모리로 인하여 저장 공간을 초과하는 문제가 발생한다. 이를 해결하기 위해 다양한 부하제한 기법들이 활발히 연구되고 있다. 부하제한 기법에는 크게 랜덤부하제한 기법과 의미적부하제한 기법, 샘플링 기법으로 분류된다. 랜덤부하제한 기법은 무작위로 데이터를 선택하여 삭제하고, 의미적부하제한 기법은 데이터의 우선순위를 부여하여 우선순위가 낮은 데이터부터 삭제한다. 샘플링 기법은 통계적인 연산을 이용하여 샘플링 비율을 산정하고 이를 토대로 부하를 제한한다. 그러나 기존 기법들은 공간적 특성을 전혀 고려하지 않기 때문에 공간 질의 정확도를 감소시키는 문제를 갖는다. 본 논문은 GeoSensor 환경에서 DSMS에 발생하는 과부하 발생을 제한하고 공간 질의 정확도를 향상시키기 위해 선-필터링을 이용한 후-부하제한 기법을 제안한다. 본 기법은 선-필터링을 통하여 스트림 큐에 불필요하게 가중되는 부하를 1차적으로 제한하며, 과부하 발생 시 공간 질의 결과 정확도를 보장하기 위하여 공간 중요도와 데이터 중요도를 고려하여 후-부하제한을 수행한다. 이 기법을 이용하여 부하제한 수행 횟수를 효과적으로 감소시켰고, 공간 질의 정확도를 향상시켰다.

키워드 : GeoSensor 환경, 데이터 스트림, 부하제한, 필터링

Abstract In u-GIS environment, GeoSensor environment requires that dynamic data captured from various sensors and static information in terms of features in 2D or 3D are fused together. GeoSensors, the core of this environment, are distributed over a wide area sporadically, and are collected in any size constantly. As a result, storage space could be exceeded because of restricted memory in DSMS. To solve this kind of problems, a lot of related studies are being researched actively. There are typically 3 different methods - Random Load Shedding, Semantic Load Shedding, and Sampling. Random Load Shedding chooses and deletes data in random. Semantic Load Shedding prioritizes data, then deletes it first which has lower priority. Sampling uses statistical operation, computes sampling rate, and sheds load. However, they are not high accuracy because traditional ones do not consider spatial characteristics. In this paper 'Pre-Filtering based Post Load Shedding' are suggested to improve the accuracy of spatial query and to restrict load shedding in DSMS. This method, at first, limits unnecessarily increased loads in stream queue with 'Pre-Filtering'. And then, it processes 'Post-Load Shedding', considering data and spatial status to guarantee the accuracy of result. The suggested method effectively reduces the number of the performance of load shedding, and improves the accuracy of spatial query.

Keywords : GeoSensor Network, Data Stream, Load Shedding, Filtering

* 본 연구는 건설교통부 첨단도시기술개발사업 - 지능형국토정보기술혁신 사업과제의 연구비지원(07국토정보C05)에 의해 수행되었습니다.

* 인하대학교 컴퓨터정보공학과 석사과정, hokim@dblab.inha.ac.kr

** 인하대학교 컴퓨터정보공학과 박사과정, {shbaek, dwlee}@dblab.inha.ac.kr

*** 서원대학교 컴퓨터교육과 부교수, gbkim@seowon.ac.kr (교신저자)

**** 인하대학교 컴퓨터정보공학과 조교수, hybae@inha.ac.kr

1. 서론

인간의 삶을 돕는 유비쿼터스 환경에서 국토에 대한 공간 및 위치 정보를 제공하는 u-GIS 공간정보 기술이 미래 유비쿼터스 환경의 핵심 기반 기술로 대두되고 있다[1]. 이런 유비쿼터스 실현의 바탕이 되는 GeoSensor 환경은 u-GIS 공간정보 기술을 필수적인 요소로 삼는다. u-GIS 공간정보 기술은 기존 GIS인 건물, 도로, 지하시설물과 같은 2차원 또는 3차원상의 정적인 지형지물 정보와 유비쿼터스 환경을 기반으로 시간에 따라 공간적인 위치가 포함된 동적인 GeoSensor 정보의 융합 처리를 요구한다[2].

여기서 GeoSensor는 넓은 지역에 산발적으로 분포하며, 고정성뿐만 아니라 자신의 위치 인식 장치를 이용하여 시간에 따라 장소를 이동할 수 있는 이동성도 가지고 있다. 넓은 지역에서 실시간으로 발생하는 대용량 정보를 처리하기 위해 데이터 스트림 처리 기술이 요구된다[2,3,4]. 즉, GeoSensor에 의해 수집되는 데이터는 데이터 스트림의 특성과 GIS의 특성을 동시에 갖는 동적인 모델이다[5].

GeoSensor에 의해 수집되는 데이터는 실시간으로 획득되며 수집되는 양은 항상 유동적이고 그 크기 역시 매우 가변적이다. 또한 다양한 이벤트에 의해 폭발적으로 데이터가 발생하는 특징을 가지고 있다. 이러한 특징은 서버의 제한된 메모리로 인하여 주어진 메모리를 초과하는 현상과 데이터가 손실되거나 유실되는 현상을 발생시킨다. 이런 현상은 결과적으로 질의 처리 결과의 정확도를 감소시키는 문제점을 야기한다[2,3,6,7,8]. 그러나 방재시스템, 교통관제시스템, 위치기반응용시스템 등과 같은 응용들은 Data Stream Management System(DSMS)에 등록된 질의를 통해 정확한 결과 값을 전달 받길 기대한다[6]. 따라서 부하를 제한하기 위한 다양한 기법들의 연구가 활발히 진행되고 있다[3,6,7,8].

부하제한 기법은 크게 랜덤부하제한 기법, 의미적부하제한 기법, 샘플링 기법으로 분류된다[3]. 랜덤부하제한 기법은 데이터의 중요도를 고려하지 않고 스트림 큐에 삽입된 데이터를 임의적으로 선택하여 삭제하는 방법으로, 무작위로 선택하기 위한 간단한 연산만을 통해 수행되기 때문에 빠른 부하제한을 수행한다. 의미적부하제한 기법은 기존의 랜덤부하제한 기법보다 정확도를 향상시키기 위해 연구된 산출물이다. 수행방식은 과부하가 발생한 스트림 큐에서 각각의 데이터의 중요도를 고려하여 우선순위를 할당한다. 할당 받은 우선순위를 통해 우선순위가 낮은 데이터는 삭제되어 우선순위가 높은 데이터가 질의 처리기에 삽입될 수 있도록 한다. 등록된 질의에 대한 결과 값의 정확도 감소를 최소화하여 결과의 신뢰성을 보장하는 장점을 갖는다[2]. 더불어 샘플링 기법은 다양한 알고리즘의 통계적인 연산을 기반으로 샘플링 비율을 산정하여 해당 비율만큼 데이터를 감소시킨다. 이 기법은 질의의 정확도를 최대한 보장하고 에러율을 균등하게 분산시키는 특징을 갖는다.

그러나 기존에 연구됐던 다양한 방법들로는 공간 연산

이 중요시 되는 GeoSensor환경에서의 DSMS에서 만족스러운 결과를 얻지 못한다. 위에서 언급했던 의미적부하제한 기법이나 샘플링 기법은 공간의 중요도를 이용한 우선순위 반영 및 비율 산정이 아니기 때문에, 특정 공간에 해당하는 데이터만이 질의 처리되는 공간 연산에 대해서는 정확도가 감소하게 된다. 또한 넓은 지역을 대상으로 하기 때문에 발생하는 데이터의 양은 폭발적일 수 있다. 이에 발생하게 되는 과부하로 인하여 질의 처리 결과의 정확도는 급격히 감소하게 된다.

본 논문에서는, GeoSensor환경에서 DSMS의 과부하 발생을 제한하고 공간 질의의 정확도 향상을 위하여 선-필터링을 이용한 후-부하제한 기법을 제안한다. 제안 기법은 이전 연구[3]를 확장한 것으로, 부하를 1차적으로 방지하기 위해 선-필터링이 선행되고 데이터가 폭발적으로 유입되어 과부하가 발생 시 후-부하제한 기법이 수행된다. 먼저, 선-필터링은 이벤트가 발생하지 않은 평상시 서버에 부하가 발생하지 않도록 데이터를 걸러내며, 이벤트 발생에 따른 데이터는 유입될 수 있도록 한다. 이를 위해서 GeoSensor로부터 유입되는 데이터의 특성을 파악하고, 스트림 큐에 가중되는 데이터 유입률을 분석한 후 주기적으로 가중치를 갱신하여 필터 범위를 결정한다. 결정된 필터 범위에 센싱 값이 포함될 경우 걸러낸다. 다음, 후-부하제한 기법은 사용자에게 의해 등록된 공간 질의의 공간 영역 이용도를 분석하고 센서에 의해 센싱된 값을 토대로 데이터의 중요도를 파악하여 절충된 중요도를 산출한다. 이 중요도를 기준으로 중요도가 낮은 데이터부터 삭제하여 부하를 제한한다.

이에 따라, 선-필터링은 스트림 큐에 가중되는 유입률을 고려하여 적응적인 필터 범위 갱신 및 필터링 수행을 통해 부하제한 기법의 수행 횟수를 효과적으로 감소시킬 수 있으며, 후-부하제한 기법은 공간적인 요소를 적극적으로 이용하여 GeoSensor환경에서의 DSMS의 공간 질의 결과의 정확도를 향상시킬 수 있다.

2. 관련 연구

2.1 GeoSensor환경

GeoSensor환경은 유비쿼터스 실현을 위해 기존 GIS인 건물, 도로, 하천, 지하시설물과 같은 정적인 지형지물 정보와 유비쿼터스 환경을 기반으로 시간에 따라 공간적인 위치가 포함된 동적인 정보를 융합 처리 할 수 있는 u-GIS DSMS를 요구한다[2]. 이 환경은 GeoSensor, GES(GeoSensor Edge Server), GSS(GeoStream Server)로 구성된다[9,10]. 각각은 다음과 같은 역할과 관계로 형성된다. GeoSensor는 RFID 리더, 모바일 RFID, 센서 노드, CCTV 등과 같이 직접 또는 간접적으로 위치를 획득할 수 있으면서, 해당 위치와 연관된 다양한 형태의 값들을 스트림의 형태로 생성하는 장치이다[1]. 넓은 지역에 산발적으로 분포하는 다수의 GeoSensor는 특정 지역에서 발생한 데이터를 수집하는 하나의 GES와 연동된다. GES는 GeoSensor로 구성된 네트워크와 연동하여 다양

한 GeoSensor를 접근하고 제어하거나 센싱 정보를 수집하는 서버이다. 다수의 GES는 공간 질의 및 다양한 연산을 수행하는 하나의 GSS와 연동된다. 마지막으로, GSS는 GES로부터 입력되는 스트림 정보에 대하여 다양한 연산자를 실행하여 의미있는 정보를 추출하거나 이벤트를 생성하고, 이를 출력 스트림으로 제공하는 서버이다. GSS는 u-GIS환경에서 요구하는 DSMS로, GES를 통해 실세계에서 지속적으로 측정되고 짧은 시간 동안 변화하는 동적인 데이터를 수신할 수 있으며, DBMS (DataBase Management System)를 통해 긴 시간 동안 변화가 발생하지 않는 정적인 데이터를 수신할 수도 있다[5,9]. 이런 데이터들은 공간 질의 및 연산 수행을 위하여, GSS의 스트림 큐에 유입되어야 한다.

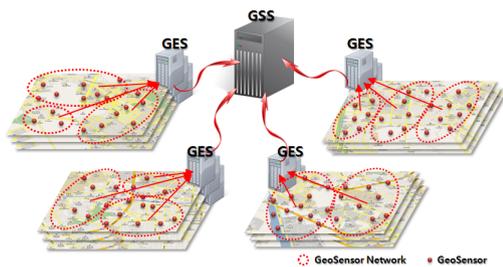


그림 1. GeoSensor환경에서의 GeoSensor, GES, GSS 간의 관계 환경도

그림 1은 GeoSensor환경에서 GeoSensor, GES, GSS 간의 관계를 나타내는 환경도이다. 이들 구성간의 공간 질의 및 데이터 흐름을 살펴보면, 사용자는 의미 있는 연산을 수행하기 위해 공간 질의를 GSS에 등록한다. GSS는 공간 질의를 분석 후, 추가적인 정보 분석을 통해 GES를 선택한다. GSS는 선택한 GES에 필요로 하는 데이터를 요청하거나 1차적인 공간 질의 수행을 요구하면, GES는 그 요청에 따라 다양한 GeoSensor로부터 데이터를 수집하며 수집된 데이터를 1차 공간 질의 수행 후 출력되는 결과 데이터들을 데이터 스트림 형태로 GSS에 전송한다. GSS로 전송된 데이터들은 사용자에 의해 등록된 공간 질의 수행에 이용되고 최종 처리된 결과는 다양한 응용에 서비스될 수 있도록 결과 데이터를 외부 상용서버로 전송한다. 이 흐름에서 보듯 GeoSensor, GES의 관계는 N:1 성립되며, GES와 GSS간의 관계 역시 N:1로 성립된다. 따라서 많은 양의 데이터를 수신하고 처리해야 하는 GES와 GSS에 적합한 부하제한을 위한 연구가 필요하다.

2.2 스트림 큐의 과부하 해결을 위한 부하제한 기법

데이터 스트림은 실시간으로 발생하며, 그 크기가 매우 유동적이다. 또한 입력 속도가 끊임없이 변화한다. 이 때문에 데이터 스트림을 처리하는 DSMS는 할당된 저장 공간이 초과하여 데이터의 손실 또는 유실로 질의 처리의 정확도를 감소시킨다. 이런 문제를 해결하기 위하여 DSMS에서는 부하제한 기법을 이용한다. 부하제한 기법은 크게 랜덤부하제한 기법과 의미적부하제한 기법, 샘플링

링 기법으로 분류된다[3].

랜덤부하제한 기법은 등록된 질의나 데이터에 대해 중요도를 고려하지 않고 무작위로 선택하여 데이터를 삭제한다. 따라서 특별하거나 복잡한 연산 과정을 생략하여 부하를 제한하는 연산 수행 속도가 빠르며, 구현 역시 간단하다. 그러나 랜덤부하제한 기법은 질의 및 데이터에 대한 중요도를 고려하지 않기 때문에, 중요하게 여기는 질의에 대해 높은 정확도를 보장하지 못할 수 있다.

의미적부하제한 기법은 랜덤부하제한 기법이 갖는 정확도를 향상시키기 위한 연구의 산출물로, 등록된 질의나 데이터의 중요도를 분석하여 우선순위를 부여하고 부여된 우선순위를 통해 데이터의 삭제 여부를 판단하게 된다. 이 기법의 수행 방식은, 스트림 큐의 데이터가 부여 받은 우선순위를 통하여 낮은 우선순위의 질의를 처리하기 위한 데이터나 데이터 자체의 우선순위가 낮은 경우 삭제한다. 이 기법은 우선순위를 통하여 사용자에게 QoS를 보장하면서, 질의 결과의 정확도를 향상시켰다. 특히 의미 있는 질의나 데이터를 선별하기 위한 기준 설정, 삭제할 데이터의 양에 따라 정확도가 크게 달라지기 때문에 우선 순위 부여 방법에 대한 연구가 활발히 진행되고 있다.

샘플링 기법은 통계적인 수치나 비율을 가지고 필요 데이터를 감소시키는 것으로, 질의의 정확도를 최대한 보장하고 에러율을 균등하게 분산시킨다. 이 기법 역시 샘플링 비율을 설정하기 위한 기준과 그 비율에 따라 질의 결과의 정확도가 달라지기 때문에 샘플링 비율을 산정하기 위한 연구가 진행 중이다. 예를 들면, Brian의 샘플링 기법은 연산마다 샘플링 비율을 가지며, 그 비율만큼 데이터를 뽑는다. 샘플링 비율을 산정하기 위하여 데이터 입력 속도와 튜플 당 평균 처리 속도 등을 고려한다[8].

GeoSensor환경에서 발생하는 데이터 스트림의 가장 중요한 요소는 공간이다. 따라서 GSS에 등록된 공간 질의를 수행하기 위해 공간 데이터 값 혹은 위치 값이 연산의 주 대상이 된다. 그러나 이와 같은 기존 부하제한 기법들은 공간적 요소를 고려하지 않고, 무작위 선택이라든가, 특정 필드의 값을 통한 우선순위 혹은 중요도 부여, 스트림 큐 내로 유입되는 속도와 데이터 분산에 의존한 샘플링 비율 산정을 이용한다. 이 때문에 공간 데이터가 존재하고 공간 질의를 수행하는 GeoSensor환경에서는 공간 질의의 처리 결과의 정확도 감소한다.

3. 선-필터링을 이용한 후-부하제한 기법

GeoSensor는 산발적으로 분포하고 수집되는 데이터는 실시간으로 획득되며 수집되는 양은 항상 유동적이고 크기 역시 매우 가변적이다. 이와 같은 특징 때문에 데이터를 최종적으로 수집하여 공간 질의를 수행하는 GSS는 데이터 스트림의 저장 공간이 초과되는 경우가 발생할 수 있다. 이런 문제를 해결하고 정확도 향상을 위하여 GES에서 수행되는 선-필터링을 이용한 후-부하제한 기법(PFPLS: Pre-Filtering Post-Load Shedding method)을 제안한다.

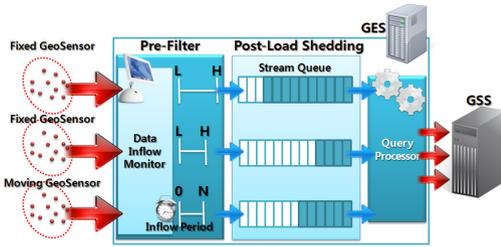


그림 2. 선-필터링을 이용한 후-부하제한 기법의 개념도

그림 2는 본 PFPLS기법의 개념도이다. 선-필터링 기법은 고정체와 이동체 GeoSensor를 구분하며, 각각의 종류에 따라 적용 방법을 달리하였다. 선-필터링 기법은 스트림 큐에 불필요하게 가중되는 부하를 제한하기 위한 목적으로 수행된다. 다음 후-부하제한 기법은 1차적으로 선-필터를 통해 부하를 제한 시켰음에도 불구하고, 스트림 큐에 과부하가 발생하게 되는 경우 수행한다. 공간 질의 검색에 의한 공간 중요도와 데이터 센싱 값에 의한 데이터 중요도를 고려하여 절충된 중요도를 산출한다. 이 절충된 중요도를 이용하여 낮은 값을 가진 데이터부터 삭제한다. 이를 통해 공간 질의 결과의 정확도와 사용자 만족도 모두를 보장할 수 있다.

3.1 적응적 필터 갱신을 이용한 선-필터링 기법

본 절에서는 GeoSensor환경에서 폭발적으로 발생한 데이터로 인하여 과부하 문제를 해결하기 위해 적응적 선-필터링(Pre-Filtering) 기법을 설명한다.

GES는 넓은 지역을 대상으로 다수의 GeoSensor로부터 많은 양의 데이터를 수집하기 때문에 부하제한 기법 적용이 필수적이다. 그러나 빈번한 부하제한은 질의 처리 결과의 정확도를 저하시키는 요인이 되기도 한다. 이전 연구[3]에서도 많은 양의 데이터가 유입되어 높은 과부하가 지속적으로 발생할 시 정확도가 급격히 감소함을 확인할 수 있었다. 더불어 각각의 스트림 큐는 데이터를 전송해주는 GeoSensor의 수, 데이터가 유입되는 속도, 유입되는 양이 다르다. 따라서 각 스트림의 데이터 유입률에 따라 적절한 가중치를 적용하여 필터링을 하는 것이 필요하다. 이를 통해 각 스트림마다 데이터의 유입률을 균등하게 유지할 수 있게 한다.

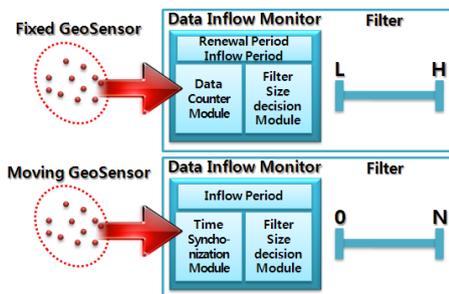


그림 3. 적응적 선-필터링 기법의 구조도

선-필터링 기법은 그림 3과 같이 고정체와 이동체 GeoSensor로 분류하여 기법 적용을 달리한다. 고정체 GeoSensor의 경우 주기적인 가중치 갱신을 통하여 필터 범위를 결정하도록 수행하며, 이동체 GeoSensor의 경우 유입 타이머를 적용하여 수행하게 된다.

우선 고정체 GeoSensor의 데이터를 전송 받는 스트림 큐에 적용되는 필터링을 설명한다. 고정체 GeoSensor의 데이터는 센서의 위치 값이 고정되어 있으며, 오로지 데이터가 발생한 시간과 해당 센서가 측정하는 센싱 값만 변화한다. 이에 센서의 측정되는 값의 변화량이 작을 경우 그 값은 질의 처리에 미치는 영향이 작기 때문에 선-필터에서 걸러낸다. 그림 2와 같이 고정체 GeoSensor로부터 데이터를 전송 받는 스트림 큐는 가중치를 달리 적용하여 필터 범위가 다르다. 필터 범위를 결정하는 가중치는 주기적으로 데이터의 유입률을 검사하여 달리 적용된다. 각 스트림 큐마다 가중치를 다르게 적용하는 이유는 각 스트림 큐마다 유입되는 데이터의 양이 다르기 때문으로 그 구체적인 환경은 다음과 같다.

- 공간 질의가 포함하는 공간 영역의 크기가 서로 다르다.
- 공간 질의가 포함하는 공간 영역 안에 존재하는 센서의 개수가 다르다.
- 따라서, 공간 질의의 공간 영역과 GeoSensor 개수 간의 관계가 다양하다.
 - 1) 공간 영역이 넓으면서 센서의 개수가 많다.
 - 2) 공간 영역이 넓으면서 센서의 개수가 적다.
 - 3) 공간 영역이 좁으면서 센서의 개수가 많다.
 - 4) 공간 영역이 좁으면서 센서의 개수가 적다.

이런 이유로, 각각의 스트림 큐에 적용되는 가중치의 주기적인 갱신을 통해 필터 범위 조절하여 스트림 큐에 유입되는 데이터의 양을 적응적으로 조절해야 한다. 만약 데이터의 유입률이 높을 경우 가중치를 높게 하여 데이터를 보다 많이 걸러낼 수 있도록 하며, 유입률이 낮을 경우 가중치를 낮게 하여 보다 적은 데이터를 걸러낼 수 있도록 한다. 이 유입률은 그림 3에서 데이터 유입 모니터링을 통하여 산정된다.

$$I^i = [N_d^i S_d / P_r] \quad (\text{식 1})$$

$$W^i = \left[I^i O(I^i) / \sum_{j=0}^C I^j \right] \quad (\text{식 2})$$

(식 1)은 유입률을 구하는 식이다, 유입률은 가중치를 결정하기 위한 가장 중요한 요소로, 유입률이 곧 스트림 큐가 받게 되는 가중치이기도 하다. i 번째 스트림 큐의 데이터 유입률 I^i 를 구하기 위해 등록된 갱신주기 P_r 동안 스트림 큐로 유입되는 데이터의 개수 N_d^i 와 데이터 크기 S_d 에 의해 결정된다.

(식 2)는 i 번째 스트림 큐의 가중치 W^i 를 구하는 식으로 갱신주기마다 수행된다. i 번째 스트림 큐의 유입률 I^i 와 GES에 생성된 하나 혹은 다수의 스트림 큐 중 해당 스트림 큐의 유입률 정렬에 의한 $O(I^i)$ 을 곱하여 유입률이 크면 클수록 높은 가중치를 받을 수 있도록 한다. 다

음, GES에 생성된 스트림 큐의 수 C_s 만큼 각 스트림 큐의 유입을 총합을 통하여 적절한 가중치를 결정한다.

$$M^i = \sum_{t=0}^{P_r} V_d^i / N_d^i, t \leq P_r \quad (\text{식 } 3)$$

$$L^i = M^i - W^i \quad (\text{식 } 4)$$

$$H^i = M^i + W^i \quad (\text{식 } 5)$$

$$T_j^i = T_d + P_j^i \quad (\text{식 } 6)$$

갱신된 가중치는 필터 범위를 결정할 수 있다. 필터 범위를 결정하기 위해서 기준 값이 산정되어 있어야 한다. 그 기준 값은 (식 3)과 같이, 갱신주기 P_r 동안 유입되는 데이터들의 값을 통해 결정되게 된다. 데이터 유입 모니터는 갱신주기 동안 i 번째 스트림 큐에 유입되는 데이터의 개수 N_d^i 를 체크하고 데이터에 포함된 센싱 값 V_d^i 을 추출하여 그 값의 평균을 구한다. 이 평균 값 M^i 는 필터 범위를 결정하는 기준 값이 된다. 구해진 필터의 기준 값은 (식 4, 5)와 같이 가중치를 적용하여 필터의 최소값 L^i 와 최대값 H^i 을 계산하여 최종적으로 필터 범위를 결정하게 된다.

선-필터링 기법은 GeoSensor의 종류와 관계없이 각각의 스트림 큐마다 유입주기를 설정해 놓는다. 이 유입주기는 사용자가 임의로 지정하며, 고정체·이동체 모두 동일한 수행 목적을 갖는다. 고정체 GeoSensor가 대상인 경우, 긴 시간 동안 GeoSensor에서 발생하는 센싱 값의 변화가 발생하지 않는다고 가정해보자. 이 데이터들은 모두 선-필터에 의해 걸러져 스트림 큐에 삽입되지 못하게 되고 GSS에서는 공간 질의를 수행해야 할 데이터가 존재하지 않게 된다. 이런 이유로 유입주기를 설정하여 해당 주기마다 필터 기능을 중지시켜 일정양의 데이터를 질의 처리될 수 있도록 한다.

다음으로 이동체 GeoSensor 데이터를 전송 받는 스트림에 적용되는 필터링을 설명한다. 이동체 GeoSensor에 의해 발생하는 데이터는 시간 값과 위치 값을 비롯하여 다양한 형태의 값들을 스트림의 형태로 가질 수 있으며, 본 논문에서는 시간 값과 위치 값에 주목하였다. 각각의 데이터는 시간 값의 변화에 따라 이동체의 위치 값이 변화한다. 따라서 같은 시간에 발생하여 서로 다른 위치 값을 가지는 데이터들이 대량으로 존재할 수 있게 된다. 예를 들어 차량에 부착된 GeoSensor가 차량 이동에 따라 위치 값을 전송한다든지, 사람이 들고 다니는 스마트폰에 부착된 GeoSensor에 의해 사람이 이동하는 위치 값을 전송하는 것이 이에 해당한다. 이 이동체 GeoSensor로부터 전송되는 데이터는 위치 값이 공간 질의 처리에 가장 중요한 요소로, 고정체 GeoSensor의 데이터와 달리 주기적으로 위치 값이 공간 질의 처리될 수 있도록 해야 한다.

그림 3에서 이동체 GeoSensor에 대한 데이터 유입 모니터를 보면 위에서 언급했던 유입주기가 존재한다. 이 유입주기를 이용해서 고정체 GeoSensor 데이터를 유입시켰던 방법과 같이 이동체 데이터의 유입 정도를 결정할 수 있게 된다. 사용자가 임의로 지정할 수 있게 하는

이유는 공간 질의 결과를 이용하는 응용에서 요구하는 정확도 정도에 따른 것으로, 매우 빈번하면서 이동 객체의 위치를 정확히 파악해야 할 경우 유입 주기를 짧게 설정하며, 이동 객체의 위치를 근사하게 파악해도 될 경우 유입 주기를 길게 설정할 수 있게 하기 위함이다.

(식 6)은 데이터의 시간 값과 유입주기를 통해 다음으로 유입될 시간을 결정하는 식이다. i 번째 스트림 큐의 유입주기 P_j^i 가 설정되면 처음 유입되는 데이터의 발생 시간 값 T_d 와 유입 주기의 동기화가 수행되면서 데이터는 스트림 큐에 삽입된다. 동기화 이후에 다음 유입 시기 T_j^i 까지 유입되는 데이터는 모두 걸러진다. 다음 유입 시기에 다다르면, 해당 시간과 일치하는 데이터는 유입함과 동시에 다시 그 다음 유입 시기를 결정한다. 이는 등록 공간 질의 수행이 종료될 때까지 반복한다.

이와 같은 선-필터링은 이벤트가 발생하지 않았음에도 일어나게 되는 과부하를 사전에 차단하여, 부하제한 기법 수행 횟수를 효과적으로 감소시킨다. 또한 스트림 큐가 갖는 데이터 유입률에 따라 주기적인 필터 범위 갱신과 유입주기를 이용하여 일정양의 데이터만을 유입시키기 때문에 스트림 큐의 메모리 사용량을 균일하게 유지할 수 있다.

3.2 공간 및 데이터의 중요도를 이용한 후-부하제한 기법

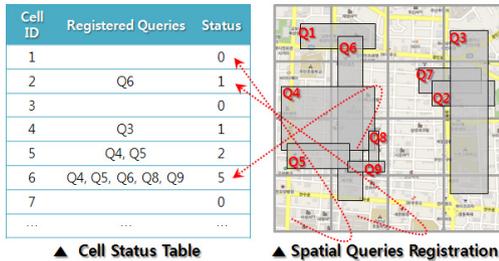
본 절에서는 스트림 큐에서 선-필터링 이후에 발생하게 되는 과부하 문제를 해결하기 위하여 공간의 중요도와 데이터의 중요도 모두를 고려한 후-부하제한 기법을 설명한다.

선-필터링 기법은 GES에 가중되는 부하를 줄임으로써 부하제한 기법 수행 횟수를 감소시키는 도구의 역할을 한다. 이를 이용하여 값의 변화가 발생하지 않는 대부분의 데이터를 걸러내지만, 넓은 지역을 대상으로 하는 GeoSensor환경에서 발생하는 데이터는 값의 변화를 예측할 수 없기 때문에 필터 범위를 벗어난 데이터가 대량으로 유입될 수 있다. 또한 다양한 이벤트 발생에 의해서도 과부하가 발생할 수 있으며 이벤트 발생에 의하여 센싱 값들의 변화가 발생하면 짧은 시간 동안 대량의 데이터들이 스트림 큐에 유입되어 결국 과부하가 발생할 우려가 있다.

이전 연구의 Different Drop 기법[3]에서는 사용자가 등록한 공간 질의들의 겹침에 따른 공간 영역의 중요도를 고려하여 부하제한을 수행한다. 스트림 큐의 과부하가 발생하면, 공간 중요도 테이블에 의해 각 데이터의 우선순위를 할당 받게 되고, 가장 낮은 우선순위를 할당 받은 데이터부터 비율적으로 삭제된다. 비율 산정은 공간 중요도의 최대 우선순위 값을 이용하며, 각각의 우선순위에 부합되도록 차별적인 비율로 구성된다. 그러나 공간 영역의 중요도만을 고려했기 때문에 사용자 기준의 정확도를 향상시켰으나, 이벤트가 발생한 데이터 기준의 중요도를 고려하지 않은 오류를 범하였다. 예를 들면, 만약 공간 중요도가 1인 영역에서 이벤트 발생하여 데이터가 유입되는 상황에서 스트림 큐에 과부하가 발생하게 되면, 이벤트가 발생했음에도 공간 중요도가 낮아 데이터가 버려

지는 문제가 발생했다. 이에 따라 공간의 중요도 및 데이터의 중요도 모듈을 고려하여 후-부하제한 기법을 수행한다. 이 기법은 GES와 GSS에서 빈번히 발생할 수 있는 과부하를 사전에 방지함과 동시에 공간 질의 수행에 있어 질의 처리 결과의 정확도를 보장할 수 있다.

본 후-부하제한 기법은 공간 중요도와 데이터 중요도를 통해 절충된 중요도를 산출하는 방식이다. 공간 중요도는 관리자에 의해 등록되는 공간 질의의 영역 겹침 정도에 비례하여 중요도가 결정되며, 데이터 중요도는 센서가 발생하는 센싱 값을 기준으로 중요도가 부여되는 형태이다.



▲ Cell Status Table ▲ Spatial Queries Registration
 그림 4. 공간 중요도 테이블 개념도

먼저, 공간 중요도를 결정하는 기준을 설명한다. 공간 중요도는 관리자에 의해 등록되는 공간 질의의 겹침 정도에 따라 결정된다. 그림 4는 공간 질의 겹침에 따라 공간 중요도 테이블에 매핑되는 모습이다. 데이터 스트림과 조인되는 공간 객체 및 공간 연산의 영역이 주어지면, 공간 질의 처리를 위해 등록된 영역을 포함하는 큰 영역을 만들고 그리드 셀로 분할한다. 그리드 셀의 분할 개수는 사용자가 임의로 설정할 수 있다. 이 그리드 셀을 이용하여 전체 영역을 균일하게 나누고, 각 셀의 등록되는 공간 질의의 수를 파악한다. 각 셀(공간 영역)의 정보를 관리하기 위하여 공간 중요도 테이블을 구축한다. 공간 중요도 테이블은 각 셀의 등록된 공간 질의의 고유 ID와 등록된 개수를 저장한다. 여기서 공간 질의의 등록 개수가 해당 셀의 중요도를 결정하는 근거가 된다. 공간 중요도를 부여하는 이유는 사용자가 원하는 값을 얻기 위하여 공간 질의를 등록하는 것이기 때문에 공간 겹침이 많을수록 사용자의 관심도 높은 공간임을 의미한다. 그림 4를 통해 예를 들면, 1번 셀의 경우 등록된 공간 질의가 없기 때문에 중요도의 값은 0이다. 반면 6번 셀과 같이 Q4, Q5, Q6, Q8, Q9가 등록된 상태이므로 해당 공간 중요도는 5로 사용자는 1번 셀보다 5번 셀의 관심이 더 높다고 판단한다. 추가로 등록되는 공간 질의가 있다면, 즉각 갱신 작업을 통하여 공간 중요도 테이블은 실시간으로 운용한다. 이 값들은 추후에 데이터 중요도와 더불어 절충된 가중치를 찾는 데 이용된다.

다음으로 데이터 중요도를 결정하는 방법을 설명한다. 데이터 중요도는 센싱되는 값을 기준으로 중요도를 부여하는 방식이다. GeoSensor의 종류는 매우 다양하다. 따라서 다양한 센서들의 종류마다 상이한 기준을 적용하여

센싱 값의 범위를 지정하여 중요도를 부여한 데이터 중요도 테이블이 구축되어야 한다. 그러나 이 테이블들을 사용하는 GES마다 개별적으로 구축한다는 것은 비효율적이다. 본 GeoSensor환경에서는 카탈로그 서버가 존재하며, 이 서버는 분산된 GeoSensor들과 다수의 GES의 정보를 추가/수정/삭제 등의 관리와 요청된 정보를 제공하는 역할을 담당한다. 데이터 중요도 테이블은 Geo-Sensor환경 상 카탈로그 서버에 등록하고 유지하는 것이 더 효율적이다.

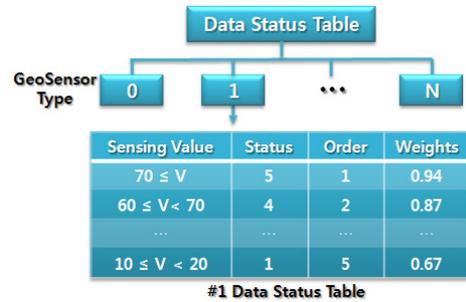


그림 5. 데이터 중요도 테이블 개념도

그림 5는 카탈로그 서버에 구축된 데이터 중요도 테이블의 모습이다. 다양한 GeoSensor들을 종류에 따라 고유번호를 부여하여 관리한다. GeoSensor의 타입과 해당 GeoSensor의 중요도테이블은 추가, 수정, 삭제될 수 있다. 데이터 중요도 테이블은 데이터의 중요도를 결정하기 위해 기준이 되는 센싱 값의 범위, 각각의 범위 항목에 따른 중요도, 중요도의 오름차순 정렬 순위 값, 각각의 중요도에 적용될 가중치로 구성된다. 이 가중치를 구하기 위하여 (식 7)을 사용한다. GeoSensor 타입 i 의 j 번째 범위 항목의 데 W_d^{ij} 이 가중치는 중요도의 총합 $\sum_{j=1}^M S_d^{ij}$ 에서 j 번째 범위 항목의 정렬 순위 값 O_d^j 을 통한 수치를 비율화하여 1을 뺀 값이다. 단, 가중치 W_d^{ij} 는 사용자에게 의해 설정될 수 있으며, 1을 넘을 수 없다.

$$W_d^{ij} = 1 - O_d^j / \sum_{j=1}^M S_d^{ij} \quad (\text{식 } 7)$$

$$S = [W_d^i \times S_d^i + (1 - W_d^i) \times S_s] \quad (\text{식 } 8)$$

지금까지 과부하가 발생한 스트림 큐에 삽입되어 있는 데이터를 대상으로 의미 있는 판단을 하기 위해 공간 중요도와 데이터 중요도를 제시했다. 이 두 가지의 중요도는 위에서 언급했듯 설정 기준이 다르다. 따라서 서로 다른 기준에 의해 산정된 중요도를 통해 절충된 의미 있는 중요도가 필요하다. 그러나 여기서 주목해야 할 점은 사용자 만족도 향상을 고려한 공간의 중요도보다 이벤트가 발생한 데이터의 중요도가 더 중요하기 때문에 가중 요소가 적용돼야 한다는 점이다. 그 이유는 과부하를 발생시킨 데이터들은 선-필터가 걸려내진 못한 데이터들로, 이들은 이벤트 발생에 의한 특별한 값이라고 간주할 수 있기 때문이다. 화재 관제 시스템을 예로 들면, 공간 중요

도가 최하인 1로 산정된 공간 영역에서 화재 발생에 의한 온도 값 상승으로 데이터 중요도가 5로 부여되었다면 어느 요소가 더 중요하게 작용해야 하는지 자명해진다.

이에 데이터 중요도의 가중치를 가중요소로 적용한 (식 8)을 사용하여 절충된 중요도를 구한다. 우선 GES는 스트림 큐의 과부하가 발생하게 되면, GES에서 자체적으로 구축하고 있는 공간 중요도 테이블과 카탈로그 서버로부터 스트림 큐로 삽입된 GeoSensor 타입과 동일한 데이터 중요도 테이블을 얻어온다. 스트림 큐의 데이터가 가지는 위치 값을 토대로 공간 중요도 테이블에서 공간 중요도 S_i 획득하고, 데이터의 센싱 값을 토대로 데이터 중요도 테이블에서 데이터 중요도 S_d^i 와 가중치 W_d^i 를 획득한다. 위에서 언급했듯 데이터 중요도의 팩터가 더 중요하기 때문에 가중치를 데이터 중요도와 곱하고 상대적으로 중요성이 낮은 공간 중요도에는 $(1 - W_d^i)$ 를 곱하여 영향을 덜 받도록 한다. 그림 5에서 제시한 데이터 중요도 테이블을 이용하여 예를 들면, 공간 중요도가 2이고 데이터의 중요도가 4인 데이터는 $[0.87*4+(1-0.87)*2]$ 으로 절충된 중요도는 4가 된다.

스트림 큐의 삽입된 데이터의 절충된 중요도가 파악되면 중요도가 낮은 데이터부터 삭제하여 부하제한을 수행한다. 결과적으로 공간 질의 검침 기준과 데이터 센싱 값의 기준으로 한 적절한 중요도를 판단하기 때문에 사용자 만족도를 보장하면서, 공간 질의 결과의 정확도를 향상시킬 수 있다.

4. 성능 분석

4.1 실험 환경

본 절에서는 제안 기법의 우수성을 분석하기 위하여 설정한 실험 환경에 대해 설명한다. 본 실험은 Fedora Core 8.0을 기반으로 수행했으며, 시스템 환경은 Intel(R) Pentium(R) 4 CPU 3.00GHz이고, 메인 메모리는 4GB이다. GES와 GSS는 Java로 구현했으며, GeoSensor에서 발생하는 데이터 셋은 센서 시뮬레이터(Sensor Simulator)로 대체하였다. 센서 시뮬레이터는 제안 기법의 성능을 평가하기 위하여 개발된 어플리케이션으로 시스템 관리 툴의 일종이다. 센서 시뮬레이터는 고정체와 이동체 GeoSensor 선택하면, 데이터의 사이즈 및 발생 주기 등 몇 가지 옵션을 설정할 수 있다. GIS 데이터는 TIGER/Line 2007 데이터로, Oracle에 구축하여 본 실험에 사용한다[15].

표 1은 성능 평가를 위한 기본 속성 값들을 정리한 표이다. 실험에 사용할 고정체 GeoSensor의 데이터는 <ID, TIME, X, Y, VALUE>로 구성되어 총 36bytes이며, 이동체 GeoSensor의 데이터는 <ID, TIME, X, Y>로 총 28bytes이다. 이 데이터들은 초당 지정된 크기만큼 데이터를 발생시킨다. 데이터가 유입되는 스트림 큐는 8Mbytes이며, 총 10개의 스트림 큐가 생성되어 있다. 10개의 스트림 큐는 서로 다른 유입률을 유지할 수 있도록 하였다. 선-필터의 기본 갱신주기는 100초 설정했으며 유

입주는 스트림 큐마다 달리 적용되기 때문에 여기서 제시하지 않는다. 후-부하제한 기법에서 사용할 데이터 중요도 테이블은 테스트를 위하여 Oracle에 임의 생성하였다. 실험 진행 시 비교 측정하는 기법은 이전 연구였던 Different Drop기법[3], Sampling 기법, Random 기법으로 선정하였다.

표 1. 실험 환경 기본 속성 값

속성	값
시스템 할당 메모리	512Mbytes
스트림 큐 사이즈	8Mbytes
스트림 큐 생성 수	10개
고정체 데이터 사이즈	36bytes
이동체 데이터 사이즈	28bytes

실험은 선-필터링을 통해 걸러지는 데이터의 양을 측정하고, 부하제한이 수행되는 횟수와 부하제한에 따른 이벤트 값이 들어간 데이터의 손실률을 타 기법과 비교 측정하며, 등록된 공간 질의 처리 결과 정확도와 사용자 만족도를 측정한다.

4.2 성능 평가

4.2.1. 데이터 필터링 되는 양 측정

본 제안 기법 중 선-필터링은 다양한 GeoSensor들로부터 유입되는 많은 양의 데이터로 인하여 발생할 수 있는 부하를 1차적으로 제한하고자 연구되었다. 따라서 선-필터링이 가지는 자체적인 성능 분석이 우선되어야 한다. 선-필터링은 고정체와 이동체 GeoSensor에 대한 적용 방법이 다르다는 것은 위에서 언급(3.1.절 참고)하였다. 각각에 대한 비교 분석을 위하여 이벤트가 발생하지 않은 고정체 센서 데이터, 이벤트가 랜덤하게 발생하는 고정체 센서 데이터, 이동체 센서 데이터를 대상으로 측정한다.

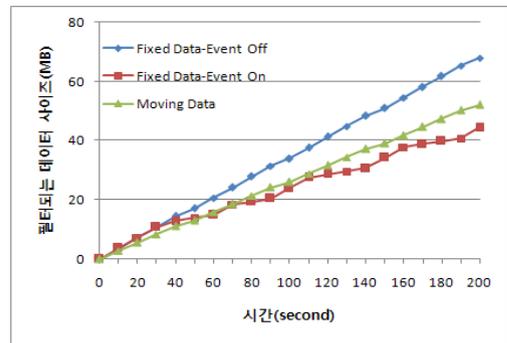


그림 6. 선-필터에 의해 걸러지는 데이터의 양 측정

그림 6은 초당 1만 건의 데이터를 발생하도록 설정하고 200초 동안 실험하여 선-필터링에 의해 걸러지는 데이터의 양을 측정된 결과이다. 이벤트가 발생하지 않은

고정체 센서 데이터의 경우 데이터의 값의 변화가 생기지 않기 때문에 모든 데이터가 걸러진다. 단, 유입주기에 의해 일부 이벤트만 스트림 큐에 삽입 되었다. 이벤트가 발생한 고정체 센서 데이터는 필터 범위를 벗어나 스트림 큐에 삽입되어 걸러진 데이터가 적다. 위 실험에서는 4번의 이벤트가 발생한 것을 확인할 수 있다. 마지막 이동체 센서 데이터에 경우, 유입주기가 50초로 설정되어 매 50초, 100초, 150초, 200초에 데이터가 삽입되었으며, 그 외에 데이터는 모두 걸러내었다.

4.2.2. 부하제한 수행 횟수 측정

이번 실험은 과부하 발생을 해결하기 위하여 부하제한이 수행되는 횟수를 측정하는 것으로, 실험을 진행하기 위하여 초당 발생하는 데이터의 수를 조정하였다. 데이터 발생 수가 증가할수록, 스트림 큐에 가중되는 부하는 더 커지며, 부하제한이 빈번하게 수행된다. 제안 기법에 경우 선-필터를 통하여 많은 양의 데이터를 걸러내기 때문에 부하제한 수행 횟수를 크게 감소시킬 수 있다.

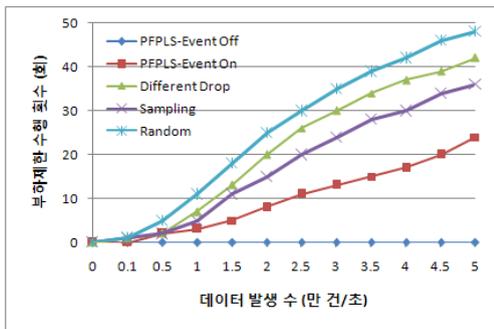


그림 7. 부하제한 수행 횟수 측정

그림 7은 초당 1천 건의 데이터 생성에서부터 5만 건의 데이터 생성에 따른 부하제한 수행 횟수를 측정한 것이다. 제안 기법(PFPLS)은 이벤트가 발생하지 않은 고정체 센서 데이터를 대상일 경우 부하제한 수행을 하지 않았다. 그 이유는 데이터의 값이 변동이 없거나 변동폭이 작아 선-필터링에 의해 모두 걸러졌기 때문이다. 더불어 이벤트가 발생하였더라도 선-필터링에 의해 1차적으로 데이터가 걸러지기 때문에 타 기법에 비해 현저하게 적은 횟수가 시행되었다.

4.2.3. 공간 질의처리 결과 정확도, 만족도, 손실률 측정

공간 데이터 스트림을 이용하는 방재시스템, 교통관제 시스템, 위치기반응용시스템 등과 같은 응용의 원활한 운용을 위해서는 정확도 보장 필수이다. 따라서 이번 실험을 통하여 제안 기법의 이벤트가 발생한 데이터의 손실률, 공간 질의 결과의 정확도, 사용자 만족도를 측정한다. 먼저, 이벤트가 발생한 데이터의 손실률 측정을 위한 실험은 초당 데이터 발생 건수를 매 실험마다 다르게 설정하여 200초 동안 진행하였다. 이벤트가 발생하는 데이터는 센서 시뮬레이터의 옵션 값을 설정하여, 200초 동안

유입되는 데이터 수 중 10%만이 발생하도록 하였으며, 유입되는 시기는 랜덤하게 설정하였다. 예를 들어, 초당 1000건이 발생하는 실험군에서 200초 동안 200,000건의 데이터가 유입되며, 이 중 20,000건의 데이터가 이벤트가 발생한 데이터로 유입되게 된다. 최종적으로 생성된 이벤트 데이터는 GSS에서 질의 처리되어 나온 결과 데이터를 대상으로 손실률을 측정하였다.

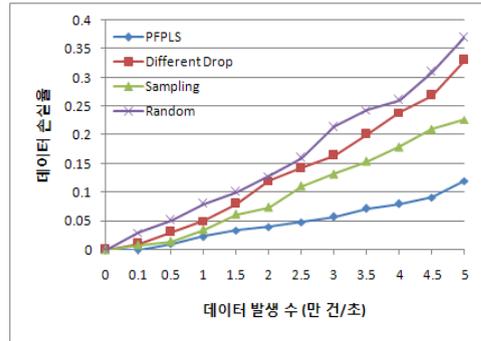


그림 8. 이벤트가 발생한 데이터의 손실률 측정

그림 8은 센서 시뮬레이터를 통해 데이터 발생 중 빈번하게 이벤트를 발생시켜 부하제한 중 손실되는 평균 비율을 측정한 결과이다. 이벤트가 발생한 데이터는 공간 질의 처리에 반영되는 중요한 데이터이며, 공간 질의 결과 정확도의 직접적 영향을 미친다. PFPLS는 데이터 중요도에 가중요소를 적용하기 때문에 이벤트가 발생한 데이터는 더 높은 중요도를 부여 받게 된다. 따라서 타 기법에 비하여 이벤트 데이터의 손실률이 낮음을 확인할 수 있다. 다른 각각의 기법들은 공간의 중요도, 샘플링 비율, 무작위 선택에 의해 데이터를 삭제하기 때문에 이벤트 데이터를 고려되지 않음을 보인다.

다음으로, 공간 질의 결과 정확도 실험 역시 200초 동안 진행하였다. 등록된 각각의 공간 질의 수행에 부합하는 데이터를 2%씩 생성되도록 하였으며, 유입되는 시기는 랜덤하게 설정하였다. 예를 들어, 초당 1000건이 발생

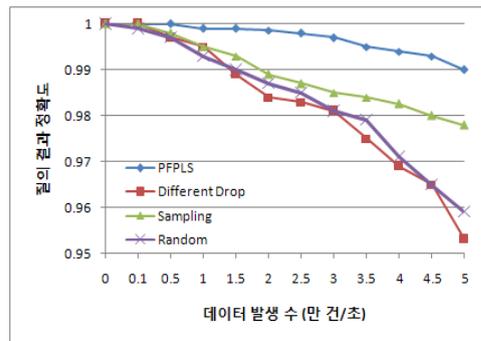


그림 9. 공간 질의 결과 정확도 측정

하는 실험군에서 200초 동안 유입되는 데이터 200,000건 중 4,000건은 특정 공간 질의 수행에 부합하는 데이터로 선정한 것이다. 즉, 15개의 공간 질의는 각각 4,000건의 데이터를 정확하게 처리해야 하며, 총 60,000건의 데이터가 질의 처리되어야 한다.

그림 9은 공간 질의 결과의 평균 정확도를 측정된 결과이다. PFPLS는 선-필터링을 통하여 부하제한 수행을 최소화하며 공간 중요도와 데이터 중요도를 통한 절충된 중요도를 고려했기 때문에 타 기법보다 높은 정확도를 보인다. Different Drop에 경우, 공간적 특성을 이용하여 부하제한을 수행하지만, 공간 질의 처리에 대상이 되는 이벤트 발생 데이터의 중요도를 고려하지 않고, 오로지 사용자가 등록한 공간 질의의 겹침에 따른 공간 중요도만을 고려했기 때문에 이와 같은 결과가 나왔다. 또한 타 기법들은 공간적 요소를 전혀 고려하지 않고 부하제한을 수행하기 때문에 본 제안 기법보다 낮은 정확도의 결과를 보인다.

이번 실험에서의 사용자 만족도를 측정하는 것으로, 3.2절에서 언급한 바와 같이 사용자에게 의해 등록되는 공간 질의의 겹침 정도가 많을수록 사용자의 관심도 높은 공간임을 의미한다. 이 실험 역시 200초 동안 진행하였으며, 7개의 공간 질의가 겹친 영역에서 10%의 데이터가 생성되도록 하였다. 예를 들면, 초당 1000건이 발생하는 실험군에서 200,000건 발생 데이터 중 20,000건의 데이터가 7개의 공간 질의가 겹친 영역에서 발생하게 된다. 최종적으로 GSS를 통해 처리된 결과 데이터 토대로 사용자 만족도를 산출하였다.

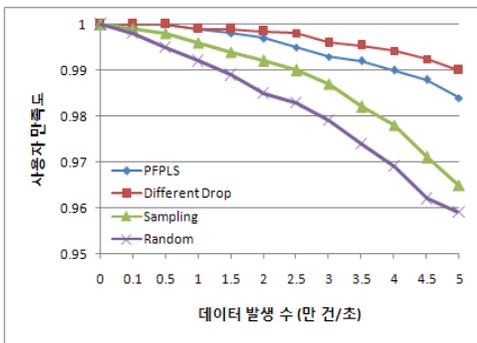


그림 10. 데이터 발생 증가에 따른 사용자 만족도 측정

그림 10은 사용자가 등록한 공간 질의의 겹침 정도에 따른 공간 중요도를 기준으로 사용자 만족도를 측정한 것이다. PFPLS는 부하제한 수행 횟수가 타 기법에 비하여 적음에도 불구하고 공간 중요도보다 데이터 중요도에 가중요소를 적용하기 때문에, 공간 중요도만을 고려하는 Different Drop보다 낮은 만족도를 보인다. 그러나 본 제안 기법은 다른 두 기법에 비해 우수한 결과를 보인다.

6. 결론

본 논문은 GeoSensor환경에서 센서 데이터를 수집하는 GES에 적용되는 선-필터링 이용한 후-부하제한 기법을 제안했다. 본 제안 기법은 선-필터링을 통하여 스트림 큐에 불필요하게 가중되는 부하를 1차적으로 제한시키며, 과부하 발생 시 공간 질의 결과 정확도를 보장하기 위하여 공간 중요도와 데이터 중요도를 고려하여 부하제한을 수행한다.

선-필터링은 고정체와 이동체 GeoSensor 데이터를 구분하며, 갱신주기에 따라 주기적으로 데이터 유입률을 고려한 후 필터 범위를 조정하여 스트림 큐에 유입되는 데이터의 양을 조정한다. 이를 통해 센싱 값의 변화가 발생하지 않거나 유입주기가 아닐 시에는 데이터를 걸러내어 부하제한 수행 횟수를 효과적으로 감소시켰다. 후-부하제한 기법은 공간 질의 겹침에 따른 공간 중요도 분석과 실제적 데이터 중요도를 고려한 절충된 중요도를 산출하여 중요도가 낮은 데이터부터 삭제한다. 공간적 특성을 충분히 반영하여 공간 질의 처리 결과의 정확도를 향상을 보였다.

향후 이동체 GeoSensor에 설정되는 유입주기의 경우 단위 시간당 이동 객체의 이동 거리를 고려하여 좀 더 정확한 필터링이 가능하도록 개선하며, 본 논문에서 제안했던 것과 같이 지속적인 선-필터링 수행이 아닌 유입률을 통해 스트림 큐에 과부하가 발생할 시점을 예측하여 해당 시점에 적절히 필터링을 수행시키는 지능형 선-필터링에 대한 연구를 진행한다. 또한 이 연구를 통해 공간 질의뿐만 아니라 비공간 질의에서도 부하제한 수행 시 정확도가 향상될 수 있도록 연구한다.

참고 문헌

- [1] 이충호, 안경환, 이문수, 김주완, "u-GIS 공간정보 기술 동향," 전자통신동향분석, ETRI, 2007.
- [2] 백성하, 이동욱, 김경배, 정원일, 배혜영, "공간 슬라이딩 윈도우 집계질의의 정확도 향상을 위한 그리드 해쉬 기반의 부하제한 기법," 한국공간정보시스템학회 논문지, 제11권 제2호, 2009, pp. 89-98.
- [3] Ho K., Sung-Ha B., Dong-Wook L., Gyoung-Bae K., Hae-Young B., "Load Shedding applying range overlap ratio of spatial query over Data Stream," ASGIS2009, 2009, pp. 49-55.
- [4] Brian B., Shivnath B., Mayur D., Rajeev M., and Jennifer W., "Models and Issues in Data Stream Systems," PODS, 2002.
- [5] Viveca A., Anita H., and Peter R., "Dynamic Information in GIS Systems," Computers Environment and Urban Systems, Vol.19 No.2, 1995, pp. 107-115.
- [6] Yi-Cheng T., Song L., Sunil P., and Bin Y., "Load Shedding in Stream Databases: A

Control-Based Approach,” In VLDB Conference, 2006, pp. 787-798.

- [7] Nesume T., Ugur C., Stan Z., Mitch C., and Michael S., “Load Shedding in a Data Stream Manager,” VLDB, 2003, pp. 309-320.
- [8] Brian B., Mayur D., and Rajeev M., “Load Shedding for Aggregation Queries over Data Stream,” ICDE, 2004, pp. 350-361.
- [9] Sang Ki K., Sung-Ha B., Dong Wook L., Warn Il C., Gyoung Bae K., Hae Young B., “Data Source Management using weight table in u-GIS DSMS,” 한국공간정보시스템학회 논문지, 제11권 제2호, 2009, pp. 27-33.
- [10] 정원일, 신승선, 백성하, 이연, 이동욱, 김경배, 이충호, 김주완, 배해영, “u-GIS 컴퓨팅을 위한 GeoSensor 데이터 스트림 처리 시스템,” 한국공간정보시스템학회 논문지, 제11권 제1호, 2009, pp. 9-16.
- [11] Xin D., Alon H., “Indexing Dataspaces,” SIGMOD, 2007.
- [12] Kittisak K., Nittaya K., “Approximate Frequent Pattern Discovery over Data Stream,” International Journal of Computer Science and Engineering, 2008, pp. 478-482.
- [13] “OpenGIS Implementation Specification for Geographic information - Simple feature access - Part1: Common Architecture,” www.opengeospatial.org, 2008.
- [14] Chris O., Jing J., and Jennifer W., “Adaptive Filters for Continuous Queries over Distributed Data Streams,” SIGMOD, 2003.
- [15] “Tiger/Line Shapefiles.”, www.census.gov/geo/www/tiger/tgrshp2007/tgrshp2007.html, 2007.



김 호
2008년 서원대학교 컴퓨터교육과 졸업 (교육학사)
2008년~현재 인하대학교 컴퓨터정보공학과 (석사과정)
관심분야는 공간 데이터베이스, 데이터 스트림, 부하제한, 질의 스케줄링



백 성 하
2005년 인하대학교 컴퓨터공학부 졸업 (이학사)
2007년 인하대학교 컴퓨터공학과 (공학석사)
2007~현재 인하대학교 컴퓨터정보공학과 (박사과정)

관심분야는 데이터 스트림, 클러스터, 위치기반 서비스



이 동 욱
2003년 상지대학교 전자계산공학과 (이학사)
2005년 인하대학교 컴퓨터공학과 (공학석사)
2005년~현재 인하대학교 컴퓨터정보공학과

(박사과정)
관심분야는 공간데이터웨어하우스, 공간정보관리, 유비쿼터스 환경을 위한 SDBMS



김 경 배
1992년 인하대학교 전자계산공학과 (공학사)
1994년 인하대학교 전자계산공학과 (공학석사)
2000년 인하대학교 전자계산공학과 (공학박사)

2000년~2004년 한국전자통신연구원 (선임연구원)
2004년~현재 서원대학교 컴퓨터교육과 부교수
관심분야는 이동 실시간 데이터베이스, 스토리지 시스템



배 해 영
1974년 인하대학교 응용물리학과 (공학사)
1978년 연세대학교 대학원 전자계산학과 (공학석사)
1989년 숭실대학교 대학원 전자계산학과 (공학박사)

1985년 Univ. of Houston 객원교수
1992년~1994년 인하대학교 전자계산소 소장
1982년~현재 인하대학교 컴퓨터공학부 교수
1999년~현재 지능형 GIS연구센터 센터장
2000년~현재 중국 중경우전대학교 대학원 명예교수
2004년~2006년 인하대학교 정보통신대학원 원장
2006년~2009년 인하대학교 대학원장
관심분야는 분산 데이터베이스, 공간 데이터베이스, 지리정보 시스템, 멀티미디어 데이터베이스