

Semantic Cloud Resource Recommendation Using Cluster Analysis in Hybrid Cloud Computing Environment

Yoonsun Ahn[†] · Yoonhee Kim^{**}

ABSTRACT

Scientists gain benefits from on-demand scalable resource provisioning, and various computing environments by using cloud computing resources for their applications. However, many cloud computing service providers offer their cloud resources according to their own policies. The descriptions of resource specification are diverse among vendors. Subsequently, it becomes difficult to find suitable cloud resources according to the characteristics of an application. Due to limited understanding of resource availability, scientists tend to choose resources used in previous experiments or over-performed resources without considering the characteristics of their applications. The need for standardized notations on diverse cloud resources without the constraints of complicated specification given by providers leads to active studies on intercloud to support interoperability in hybrid cloud environments. However, projects related to intercloud studies are limited as they are short of expertise in application characteristics. We define an intercloud resource classification and propose semantic resource recommendation based on statistical analysis to provide semantic cloud resource services for an application in hybrid cloud computing environments. The scheme proves benefits on resource availability and cost-efficiency with choosing semantically similar cloud resources using cluster analysis while considering application characteristics.

Keywords : Application Characteristics, Hybrid Cloud Computing, Intercloud, Semantic Cloud

군집분석을 이용한 하이브리드 클라우드 컴퓨팅 환경에서의 시맨틱 클라우드 자원 추천 서비스 기법

안 윤 선[†] · 김 윤 희^{**}

요 약

하이브리드 클라우드 컴퓨팅 환경에서 많은 과학자들이 과학 응용을 수행하고 있으나, 클라우드 컴퓨팅 서비스를 제공하는 각 회사들의 자원 표기법이 상이하고 복잡하여 사용에 어려움이 따르고, 응용에 적합한 클라우드 자원을 선택하는 것이 어렵다. 클라우드 서비스 간에 상호 호환성을 제공해주는 하이브리드 클라우드 환경에서의 표준화된 자원 명세 표기법이 필요하다. 과학자들은 기존에 자신들이 수행했던 자원이나 가장 좋은 성능의 자원에서만 수행하려는 경향이 있어, 비용, 시간을 효율적으로 수행하면서 응용에 적합하고, 기존의 실험과 유사하게 진행할 수 있는 자원을 추천해주는 서비스가 필요하다. 하이브리드 클라우드 서비스의 표준화를 위해 인터클라우드 프로젝트가 진행되고 있으나, 과학 응용 실험에 적합한 자원의 선택을 위해 필요한 클라우드 자원의 특성들을 나타내는 데 한계가 있다. 본 논문에서는 하이브리드 클라우드 환경에서 시맨틱 클라우드 자원 서비스를 제안한다. 통계 기법으로 과학 응용의 특징에 따라 응용에 적합한 클라우드 자원을 그룹으로 분류하고 분류된 유사한 클라우드 자원 그룹을 가지고 시맨틱 클라우드 자원 추천 서비스 기법을 제공한다. 제안한 알고리즘을 통해 시맨틱 클라우드 추천 서비스 기법을 제공하면, 효율적인 자원의 가용성과 비용으로 응용을 수행할 수 있고, 응용에 적합한 클라우드 자원을 추천할 수 있다.

키워드 : 응용 특성, 하이브리드 클라우드, 인터클라우드, 시맨틱 기법

1. 서 론

클라우드 컴퓨팅은 장소나 시간에 제약 없이 IT 자원을 필요한 만큼 빌려 쓸 수 있는 방식으로 확장성과 접근성의 용이함이 있으며, 다양한 종류의 클라우드 자원을 필요한 만큼 빌려 쓸 수 있어 과학 응용 실험을 수행하는 과학자들이 늘고 있다.

※ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2013R1A1A3007866).

† 준 회 원 : 숙명여자대학교 컴퓨터과학부 석사

** 종 신 회 원 : 숙명여자대학교 컴퓨터과학부 교수

Manuscript Received : July 6, 2015

First Revision : September 1, 2015

Accepted : September 1, 2015

* Corresponding Author : Yoonhee Kim(yulan@sookmyung.ac.kr)

하이브리드 클라우드 컴퓨팅 환경에서 많은 과학자들이 과학 응용을 수행하고 있으나, 다양한 클라우드 자원들의 활용과 응용에 적합한 자원을 찾는 것은 어렵다. 또한, 과학자들은 기존에 자신들이 수행했던 자원이나 가장 좋은 성능의 자원에서만 수행하려는 경향이 있어, 비용, 시간을 절약하면서 응용에 적합하고, 기존의 실험과 유사하게 진행할 수 있는 자원을 추천해주는 서비스가 필요하다.

클라우드 컴퓨팅 서비스를 제공하는 회사들은 Amazon EC2[1], Microsoft Window Azure[2], Rackspace[3], KT Ucloud[4], SCALR[5], RightScale[6], 이노그리드 Cloudit[7] 등과 같이 다양하여 사용자들이 선택할 수 있는 자원이 많이 존재한다. 그렇지만 각 회사별로 제공하는 자원의 표기법이 상이하여 응용에 적합한 자원을 찾는 것은 어렵고, 복잡하다. 그래서 클라우드 서비스 간에 상호 호환성이 필요하다.

하이브리드 클라우드 환경에서 사용자가 클라우드 컴퓨팅 서비스 제공 회사들과 관계없이 자원을 사용할 수 있게 하기 위해 자원 명세의 표준화된 표기법이 필요하다.

클라우드 서비스의 표준화된 표기법을 만들기 위해 인터클라우드 프로젝트가 진행되고 있다[8]. mOSAIC 프로젝트에서는 클라우드 자원에 대한 명세를 온톨로지를 기반으로 표현하여 자원의 공통된 특성을 개념화하고 의미 있게 그들 사이의 관계를 나타낼 수 있도록 하여 서로 다른 제공자 간의 상호 호환성을 제공하는 mOSAIC 온톨로지를 구축하였다[9]. 그러나 과학 응용 실험에 적합하고 효율적인 자원의 추천을 위해 필요한 클라우드 자원의 특성을 모두 고려하고 있지 않다.

본 논문에서는 하이브리드 클라우드 환경에서 응용의 특성에 적합한 클라우드 자원을 분류하고 추천해주는 시맨틱 클라우드 자원 추천 서비스 기법을 제안한다. mOSAIC 프로젝트[9]의 mOSAIC 온톨로지를 활용 및 확장하여 자원 명세를 정의하고, 자원 명세에 따라 클라우드 컴퓨팅 서비스 제공 회사들의 다양한 자원들을 하이브리드 클라우드 환경에서 제공 회사에 상관없이 클라우드 자원을 사용할 수 있도록 공통적인 요소들을 가지고 나타낸다. 통계적 기법인 군집 분석[10]을 기반으로 응용의 특징을 고려하여 응용에 적합한 시맨틱 클라우드 자원 그룹으로 분류한다. 응용의 특징에 따라 자원의 특성 중 가중치를 부여할 요소들 다르게 하여 응용에 적합한 다양한 클라우드 자원 그룹을 형성한다. 분류한 그룹 내에 유사한 자원들을 가지고 시맨틱 클라우드 서비스를 제공한다.

본 논문의 구성은 다음과 같다. 1절은 서론에 대해서 설명하고, 2절은 관련 연구에 대해서 설명한다. 3절에서는 시맨틱 클라우드 자원 추천 서비스 기법에 대해 설명한다. 4절에서는 제안하는 시맨틱 엔진 알고리즘에 대해서 설명한다. 5절에서는 제안한 알고리즘을 바탕으로 실험을 진행하고, 결과를 분석한다. 마지막으로 6절에서 본 논문의 결론을 맺는다.

2. 관련 연구

효율적인 하이브리드 클라우드 환경이 되기 위해서는 클라우드 간의 상호 호환성이 필요하다. 클라우드 서비스를

제공하는 자원의 특성 명세는 상이하여 이를 해결하고자, 다양한 연구가 진행되고 있다.

서로 다른 자원의 특성 명세 문제를 해결하기 위한 표준화 프로젝트인 인터클라우드 프로젝트가 진행되고 있어, 많은 연구에서 인터클라우드 표준화 프로젝트에서 진행된 내용을 가지고 연구를 수행하고 있다[11]. 인터클라우드 표준화 프로젝트 중, mOSAIC 프로젝트[9]는 다양한 클라우드 서비스 사이의 상호 호환성을 제공하여 이기종 클라우드 컴퓨팅 자원에 쉽게 접근하고, 간단하게 사용할 수 있게 하며, SLA를 제공하고자 한다. mOSAIC 프로젝트[9]는 클라우드 컴퓨팅의 서비스들을 정리하여 제공하기 위해서 여러 클라우드 분류법들을 기반으로 클라우드 온톨로지를 구축하였다.

mOSAIC 프로젝트[9]에서는 클라우드 온톨로지인, mOSAIC 온톨로지를 구축하여 클라우드 자원의 공통된 특성을 개념화하고, 상호 호환성을 제공해주며, 시맨틱 검색과 클라우드 서비스의 통합 환경을 제공해주고 있다. 그러나 클라우드 자원 제공자들이 제공하는 자원의 특성을 모두 고려하고 있지 않다.

시맨틱 클라우드 온톨로지를 기반으로 클라우드 간의 의미적 계층 관계를 정의하여 사용자의 요구사항에 따라 검색의 효율성을 향상시킬 수 있다. 온톨로지 기반 시맨틱 클라우드의 태그분류는 의미상의 계층 관계를 정의한 트리구조로 구성되어 있다. 시맨틱 클라우드는 시맨틱 기반 계층 관계를 고려한 형태의 클라우드라고 볼 수 있다[12].

클라우드 컴퓨팅 환경에서 시맨틱 기술을 바탕으로 클라우드 자원 관리 기법 연구들이 진행되고 있다. 시맨틱 기술은 온톨로지를 이용하는 방식과 다양한 통계적 기법을 활용한 방식들로 분류할 수 있다.

온톨로지는 해당 도메인에서 용어와 개념 혹은 의미를 관계와 함께 묘사하여 도메인의 지식을 표현할 수 있는 대표적인 시맨틱 기술의 하나로 클라우드 환경에서 온톨로지를 기반으로 한 연구들이 다양하게 진행 중이다[16-19].

Rodriguez-Garcia[13]는 자동적인 시맨틱 어노테이션(Semantic Annotation) 플랫폼을 제공하고 있다. 클라우드 서비스 명세들로부터 의미론적인 내용들을 모으고, 자연어를 입력값으로 하는 의미론적인 검색 엔진을 제공하기 위해서 온톨로지를 구축하고, 유사도를 계산하여 서비스를 제공하였다. 여러 도메인을 처리하기 위해 멀티-온톨로지 환경을 고려한 플랫폼을 제공하여 클라우드 서비스를 제공하였다. 그렇지만 사용자가 원하는 내용으로 검색은 불가능하다.

Li Liu[14]는 클라우드 서비스를 찾아주는 것의 정확도를 높이는 것에 중점을 두고 있다. 온톨로지를 기반으로 클라우드 서비스를 제공해주고 있어, 온톨로지를 이용하여 두 개념들 간에 정확하게 같지 않아도 유사도를 계산하여 검색 서비스를 제공한다.

시맨틱 기술로 다양한 통계적 기법을 활용한 연구들이 수행되고 있다. Qingtao Wu[15]는 클라우드 서비스 자원을 베이저안 분류 알고리즘(Bayesian Classification Algorithm)을 이용하여 분류하고 있다. 자원을 여러 가지 요소들의 유사성 계산으로 분류하고 있으며, 특히 어떤 요소들은 다른 요소들보다 분류에 영향을 크게 주기 위해서 가중치를 고려한

베이지안 식을 이용하여 특징의 유사성을 계산하였다. 클라우드 서비스 자원의 분류까지 고려하고 있으며 분류한 내용을 바탕으로 서비스는 제공하지 않고 있다.

본 논문에서는 하이브리드 클라우드 환경에서의 자원의 상호 호환성을 제공하기 위해 자원 명세에 따라 클라우드 자원을 정리하고, 통계적 기법 중 군집 분석을 기반으로 유사한 클라우드 자원을 추천해주는 시맨틱 클라우드 서비스를 제공한다.

3. 시맨틱 클라우드 자원 추천 서비스 기법

효율적인 클라우드 자원 사용을 위해 하이브리드 클라우드 환경에서의 시맨틱 클라우드 자원 추천 기법을 제안한다. 제안된 기법은 하이브리드 클라우드 컴퓨팅 환경에서의 클라우드 자원의 특성을 명세하고, 군집 분석을 기반으로 시맨틱 클라우드 자원 추천 서비스 기법을 수행한다.

하이브리드 클라우드 환경에서 클라우드 제공자들의 자원들을 쉽고 간편하게 사용하고자 클라우드 자원의 특성을 명세한다. 다양한 클라우드 서비스 제공자들의 상이한 표기법을 통합하기 위해 필요한 자원의 특성을 정의하였다.

mOSAIC 프로젝트[9]의 mOSAIC 온톨로지를 활용 및 확장하여 다양한 클라우드 서비스 간에 상호 호환성을 제공할 수 있는 자원의 특성을 명세하였다. 자원의 특성 명세는 9개의 요소들로 구성되어 있다. mOSAIC 온톨로지서 제공하는 요소들 중 일부와 추가로 vCPU, Cost 부분을 요소로 사용하였다. 클라우드 자원 제공 회사명과 자원의 이름을 나타내는 요소들이 있고, 실제로 가상 클라우드 자원이 수행되는 물리 머신에 대한 정보들인 NetworkBandwidth, MemorySize, CPU Flops 정보, CPU core 개수가 나타난다. Component 부분에는 가상 클라우드 자원의 특성을 나타내는 vCPU, Cost, Storage 부분이 있다.

자원의 특성 명세에 따라 클라우드 컴퓨팅 서비스 제공 회사들의 다양한 자원들의 특성을 공통적인 요소들을 가지고 나타내어 하이브리드 클라우드 컴퓨팅 환경에서 자원 제공 회사에 상관없이 클라우드 자원을 선택할 수 있다.

공용 클라우드 자원 제공자(Amazon EC2[1], Microsoft Window Azure[2], KT Ucloud[4], rackspace[3])의 클라우드 가상 자원과 사설 클라우드인 Openstack[16]의 클라우드 가상 자원을 자원 특성 정의를 적용하여 명세하였다. Amazon EC2[1]에서 제공하는 41개 자원, Microsoft Window Azure[2]의 20개 자원, KT Ucloud[4]의 20개 자원, rackspace[3]의 19개, Openstack[16]의 7개 자원을 명세하여 여러 클라우드 자원 제공자들의 107개의 클라우드 가상 자원을 자원의 특성 명세에 따라 정리하여 시맨틱 클라우드 자원 추천 서비스 기법에서 수행할 수 있도록 하였다.

응용에 적합한 자원을 추천하기 위해서 응용에 대한 정보인 응용 명세가 필요하다. 어떤 특징을 가지는 응용인지를 파악하여 응용의 특징에 따라 자원의 요소를 고려하여 적합한 자원을 추천하게 된다. 응용 특성 명세는 이전 연구

[17]에서 수행했던 내용을 바탕으로 응용 특성을 참고하여 적합한 자원을 찾기 위해 가중치를 고려할 자원의 요소를 선택하였다.

시맨틱 클라우드 자원을 추천해주는 서비스를 하이브리드 클라우드 환경에서 제공한다. 추천 서비스를 제공하기 위해서 시맨틱 엔진은 클라우드 자원과 응용에 대한 내용을 바탕으로 클라우드 자원들 간의 유사성을 계산하고 분석한다.

4. 시맨틱 엔진 알고리즘

시맨틱 엔진은 시맨틱 클라우드 자원 추천 서비스들이 수행되기 위한 기법이다. 시맨틱 엔진이 수행되기 위해서 필요한 클라우드 자원 정보는 클라우드 가상 자원을 제공하는 제공자들의 공용 클라우드 자원과 개인적으로 구축하여 사용하는 사설 클라우드 자원 모두를 고려한, 하이브리드 클라우드 자원들을 클라우드 자원의 특성 명세를 바탕으로 자원의 명세를 동일하게 표현하여 클라우드 자원 간의 유사성 계산이 가능하도록 시맨틱 엔진에 제공한다. 또한, 응용의 특성에 맞는 시맨틱 클라우드 자원 추천 서비스를 제공하기 위해 응용의 특징을 시맨틱 엔진에 제공한다.

시맨틱 엔진은 시맨틱 클라우드 자원의 그룹을 형성하기 위해서 군집 분석을 이용하여 클라우드 자원들의 유사성을 계산하여 그룹으로 분류한다. 군집 분석은 데이터를 유사한 특성의 그룹으로 나누는 데 유용한 통계적 기법이다. 군집 분석에서 가중치를 고려하게 되면 가중치에 따라 다양한 그룹을 만들 수 있어, 응용의 특징에 따라 가중치를 고려하여 응용의 특징에 적합한 그룹을 형성하게 된다.

시맨틱 엔진이 유사성 계산을 통해 만든 그룹 내의 자원들은 의미론적으로 유사한 자원들을 나타낸다. 정확히 일치하지 않는 클라우드 자원이라도 의미상 유사한 자원을 통합하고, 각 그룹별로 그룹을 대표하는 대표 자원을 선택한다. 선택된 대표 자원을 바탕으로 자원을 추천한다.

군집 분석 기반 시맨틱 클라우드 자원 분류 알고리즘의 흐름은 Fig. 1과 같다. 본 알고리즘은 시맨틱 클라우드 자원 추천 서비스 기법에서 시맨틱 엔진 부분에서 수행된다.

클라우드 자원 정보와 응용 명세를 바탕으로, 응용의 특성에 따라 가중치를 설정하여 클라우드 자원 간의 유사성을 계산하고, 그룹으로 분류하게 된다.

시맨틱 클라우드 자원으로 분류하기 위해서 자원 특성 명세에 따라 정리된 클라우드 자원들을 두 개씩 비교한다. 두 자원의 각각의 특성을 비교하여 두 자원의 유사도를 계산한다. n개의 특성을 하나하나 비교하여 모든 특성 비교값의 합으로 두 자원의 유사도를 계산한다. 두 자원의 특성을 계산할 때, 특성이 같으면 1, 다르면 0으로 표현한다.

두 자원의 특성이 유사하면 유사도가 큰 값이 된다. 이때, 응용의 특성을 고려하는 경우, 가중치를 고려하여 계산하게 된다. 응용특성에 따라 선택된 특성에 가중치가 주어진다면, 두 자원의 특성의 유사성을 계산할 때, 같으면 상수값으로, 다르면 0으로 표현한다.

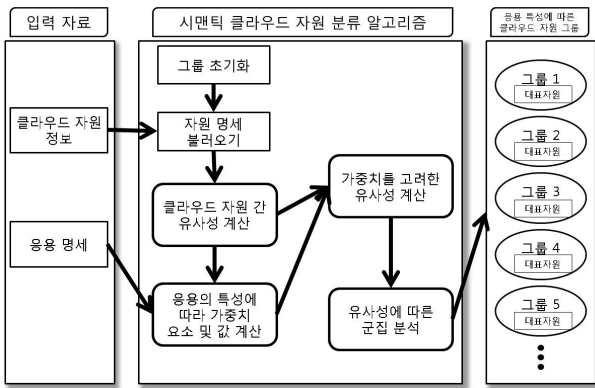


Fig. 1. Semantic Cloud Resource Analysis Algorithm

가중치를 고려하여 유사도를 계산하여 응용의 특징에 맞는 자원의 그룹을 형성할 수 있다. 특정 응용에서 중요하다고 판단되는 자원의 특징에 가중치를 주어 자원 간의 유사성을 비교하면, 가중치가 높은 자원의 특징을 유사하게 가진 자원들이 더 큰 유사성을 가지게 된다. 모든 자원을 각각 비교하여 행렬 형태로 나타낼 수 있다. 유사도값을 계산하여, 자원들 간의 유사도값을 행렬로 나타낸다. 행렬로 나타낸 것을 이용하여 유사도가 큰 자원끼리 묶어서 그룹으로 만들게 된다. 가장 유사도값이 큰 자원들을 선택하여 같은 그룹으로 묶고, 그룹이 형성된 이후에 새로이 행렬을 만들게 된다. 그리고 새롭게 생성된 행렬에서 최댓값을 찾아 그룹으로 묶어주고 또다시 그룹으로 묶기 위해 새로이 행렬을 만드는 작업을 반복하면, 응용 특성에 따른 그룹이 형성된다.

한 그룹 내의 자원들은 의미론적으로 유사한 자원들을 나타낸다. 군집 분석을 통해 하나의 그룹 내에 있는 자원들은 유사한 자원이라는 판단이 가능하여, 유사 자원을 추천할 때, 같은 그룹 내에 있는 자원을 찾아주게 된다.

시맨틱 자원 그룹으로 형성한 후, 각 그룹 내에 대표 자원을 선택한다. 각 특성들의 평균값을 계산하여 모든 특성들의 평균과 가장 유사한 것을 찾아 대표 자원으로 선택한다.

같은 그룹 내에 있는 자원으로 같은 응용을 수행하였을 때, 응용 수행 결과는 유사하게 나타난다. 정확히 일치하지 않는 클라우드 자원이라도 의미상 일치하는 자원을 통합하여 관리, 선택하는 것이 가능하다.

5. 실험

제안한 군집 분석 기반 시맨틱 클라우드 자원 선택 모델을 기반으로 클라우드심[18]을 이용하여 실험을 진행하고 그 결과를 분석하여 제안한 방법의 성능을 검증한다. Bag of Tasks[19] 형식의 전산 유체 역학 응용(난류 유동 수치해석)을 제안한 알고리즘에 적용하였다. 자원의 특성 명세를 바탕으로 정리된 하이브리드 클라우드 자원들을 시맨틱 자원들의 그룹으로 형성하고, 시맨틱 자원 그룹들을 가지고 실험을 진행한다.

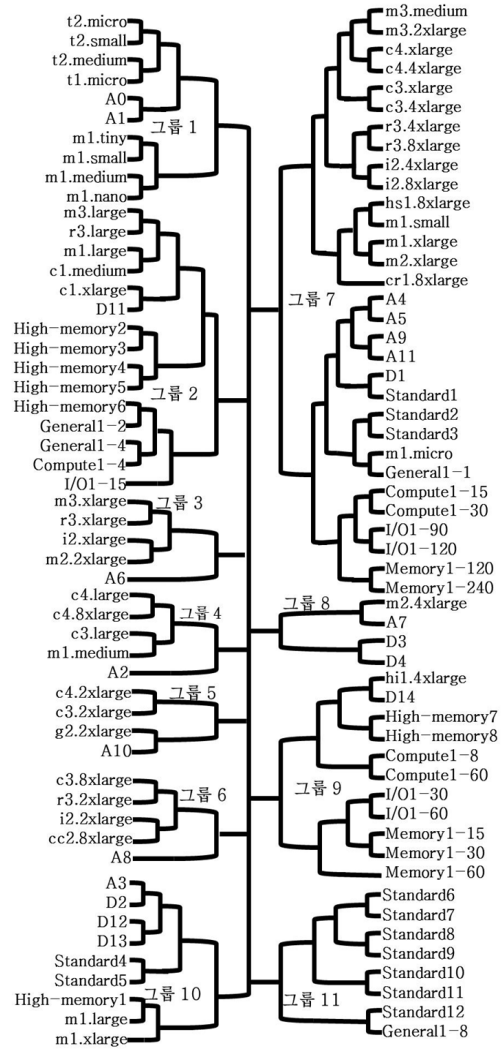


Fig. 2. Result of Grouping Cloud Resources

전산 유체 역학[20]은 유체 현상을 수치 기법의 알고리즘을 사용하여 유체 유동 문제에서 유체와 기체의 상호작용을 시뮬레이션 하는 것이다. 컴퓨터 계산을 이용하는 유체유동 해석 분야 중 난류 문제를 풀고, 적절한 난류 모델을 도입하여 문제를 단순화한 후 수치 시뮬레이션을 수행하는 것이 필요하며 이를 난류 유동 수치 해석 응용이라고 한다. 난류 유동 수치 해석 응용은 시뮬레이션에서 가장 작은 스케일에 대한 가장 큰 스케일의 비율은 격자수와 관련이 있으며, 격자수가 증가할수록 계산량이 증가하게 되며 작업 수행 시간이 길어진다. 컴퓨터의 속도 및 메모리 용량에 따라서 해석이 가능한 최대 레이놀즈수가 결정되어 계산 처리량이 결정된다. 또한, 응용 수행 전에 필요한 설치 파일들이 존재한다. 그래서 전산 유체 역학 응용인 난류 유동 수치 해석 응용의 특성에 따라 Memorysize와 vCPU, Storage, Cost를 가중치로 고려하였다. Fig. 2는 난류 유동 수치 해석 응용의 특성을 고려하여 유사한 시맨틱 클라우드 자원 그룹으로 나타난 결과이며, 11개의 그룹으로 분류되었다.

가중치를 고려하는 것에 따라서 그룹이 다양하게 형성되고, 응용의 특성에 맞는 그룹이 형성되어, 그룹들을 바탕으로 응용에 적합한 시맨틱 클라우드 자원 추천 서비스를 제공해주게 된다.

Amazon EC2[1] 4개의 자원(t2.micro, m3.2xlarge, c4.xlarge, r3.large)과 Microsoft Window Azure[2]의 5개 자원(A2, A4, A6, D2, D13) 및 Openstack[16] 3개 자원(m1.small, m1.medium, m1.large)에서 난류 유동 수치 해석 응용을 수행하여 작업의 평균치를 계산하여 클라우드스킴[18]에서 시뮬레이션을 수행하였다.

5.1 실험 1 - 시맨틱 자원 추천 서비스 비교 분석

과학자들은 실험을 진행할 때, 자신들이 이전 실험에서 실행했던 자원에서만 응용을 수행하려고 하거나 가장 성능이 좋은 자원에서 응용을 수행하려는 경향이 있다. 비용이나 시간을 절약하면서, 이전 실험의 결과와 유사하게 진행할 수 있는 자원들을 선택하는 서비스를 제공해주는 것이 가능하다는 것을 보이기 위해 실험 1을 진행하였다.

실험 1의 경우, 10,000개 작업을 성능이 가장 좋은 자원들과 난류 유동 수치 해석 응용의 특성을 고려하여 비용이 저렴하면서 실험 결과와 성능이 가장 좋은 자원들과 유사한 시맨틱 자원들의 그룹인 그룹 9의 자원들을 가지고 수행한 결과를 비교·분석하였다.

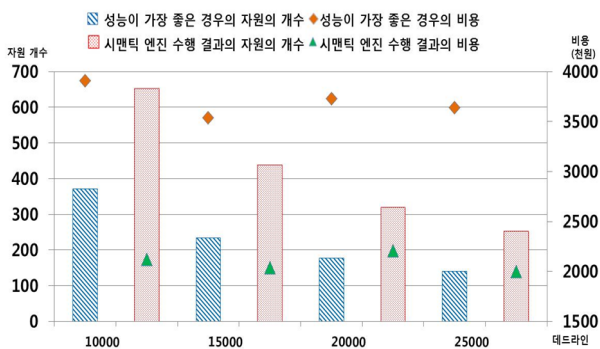


Fig. 3. Result of Semantic Cloud Resource Service

성능이 좋은 자원들을 사용하면 자원의 개수는 적으나 비용은 많이 든다. 반면, 시맨틱 자원들 중에서 비용이 저렴한 자원을 선택한 경우, 성능이 떨어져서 사용하는 자원의 개수는 증가하나 테드라인 내에 모든 작업이 완료되고, 비용도 절감된다.

I/O1-120(Microsoft Window Azure[2]) 자원은 성능이 가장 좋으나 가격은 매우 비싸다. 그래서 I/O1-120(Microsoft Window Azure[2]) 자원을 가지고 수행하면, 자원 개수는 줄일 수 있으나 가격이 매우 비싼 것을 알 수 있다.

반면, High-memory7(KT Ucloud[4]) 자원의 경우 성능은 I/O1-120(Microsoft Window Azure[2]) 자원보다 조금 떨어지지만, 비용이 저렴하다. 제안한 알고리즘으로 High-memory7(KT Ucloud[4]) 자원이 속한 그룹 9의 자원들로 작업을 수행하면, 테드라인 내에서 자원의 개수는 늘어나지만, 비용은

절감된다. Fig. 3에서 테드라인이 10,000초일 때, 성능이 가장 좋은 자원들의 사용 개수는 372개이나, 비용은 3,906,000원이다. 그러나 시맨틱 엔진의 결과인 시맨틱 자원 중 저렴한 가격의 자원을 선택한 경우는 사용 자원의 개수는 653개이고, 비용은 2,121,600원으로 약 2배의 자원을 더 사용하였으나, 비용은 약 45% 절감한 것을 볼 수 있다.

제안하는 시맨틱 클라우드 추천 서비스 기법은 성능이 가장 좋은 자원을 사용했을 경우와 수행 결과는 유사하나 비용을 절감할 수 있는 자원을 추천하는 서비스를 제공한다.

5.2 실험 2 - 난류 유동 수치 해석 응용의 특성에 따른 수행 결과 비교

실험 2는 10,000개 작업의 난류 유동 수치 해석 응용을 가중치를 고려하지 않고 나타난 시맨틱 자원들에서 수행한 결과와 난류 유동 수치 해석 응용의 특성을 고려하여 형성된 시맨틱 자원을 가지고 수행한 결과를 비교하였다.

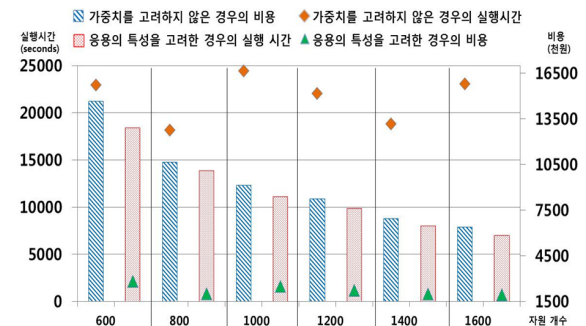


Fig. 4. Result of Semantic Engine Using Application Characteristics

실험 2는 10,000개 작업의 난류 유동 수치 해석 응용을 가중치를 고려하지 않고 나타난 시맨틱 자원들에서 수행한 결과와 난류 유동 수치 해석 응용의 특성을 고려하여 형성된 시맨틱 자원을 가지고 수행한 결과를 비교하였다.

가중치를 고려하지 않은 경우는 가중치를 고려하지 않고 형성된 자원 그룹 중에 하나를 선택하여 수행한 결과이며, 응용의 특성을 고려한 경우는 Memorysize와 vCPU, Storage, Cost에 가중치를 고려하여 형성한 Fig. 2에서 그룹 5에 있는 자원들을 가지고 수행한 결과이다.

응용의 특성을 고려했을 때 가중치를 고려하지 않고 수행한 경우보다 실행시간이 빨라졌고, 비용도 절감하여 작업을 수행하였다. 자원의 개수가 600개인 경우, 가중치를 고려하지 않은 경우보다 실행시간은 2,800초 단축하여, 13% 감소하였으며, 자원의 개수가 1,400개일 때, 응용의 특성을 고려한 경우, 약 85%의 비용을 절감하였다. 응용의 특성을 고려하여 비용은 약 85%가 절감되었고, 실행시간은 약 9.8% 단축되어 난류 유동 수치 해석 응용의 계산 처리량을 높였고, 비용을 줄여 응용에 적합한 자원들을 선택하였다.

이처럼, 군집 분석 기반 시맨틱 클라우드 자원 추천 서비스 기법이 응용의 특성을 고려하여 응용에 적합한 시맨틱 자원들을 선택하여 효율적인 작업의 수행이 가능하다.

6. 결 론

본 논문에서는 하이브리드 클라우드 환경에서 시맨틱 자원을 분류해주는 시맨틱 클라우드 자원 추천 서비스 기법을 제안하였다. 통계적 기법인 군집 분석으로 응용의 특징을 고려하여 응용에 적합한 시맨틱 클라우드 자원으로 분류하고, 대표 자원을 선택하여 시맨틱 클라우드 자원을 추천한다.

제안한 알고리즘을 기반으로 실험을 수행한 결과 시맨틱 자원 그룹 내의 자원들의 수행 결과는 유사하고, 응용의 특성을 고려하여 응용에 적합한 클라우드 자원 추천이 가능하다는 것을 보였다.

제안한 시맨틱 클라우드 자원 추천 서비스 기법은 비용이나 시간을 절약하면서, 기존에 수행한 실험과 유사하게 실험을 진행할 수 있도록 자원을 추천하는 것과 효율적인 응용 수행을 위해 응용에 적합한 클라우드 자원 추천 기능을 제공한다.

향후에는 클라우드 자원 환경을 추가, 확장하여 본 논문에서 제안한 알고리즘을 보완하여 적용할 것이다.

References

[1] Amazon EC2 [Internet], <http://aws.amazon.com/>.
 [2] Microsoft Window Azure [Internet], <http://www.window.azure.com/>.
 [3] rackspace [Internet], <http://www.rackspace.com/>.
 [4] KT Ucloud [Internet], <http://ucloudbiz.olleh.com/>.
 [5] SCALR [Internet], <http://scalr.com/>.
 [6] RightScale [Internet], www.rightscale.com/.
 [7] Cloudit [Internet], www.cloudit.co.kr/.
 [8] Intercloud project [Internet], <http://www.intercloudtestbed.org/>.
 [9] Moscato, F., Aversa, R., Di Martino, B., Fortis, T., Munteanu, V., "An Analysis of mOSAIC ontology for Cloud Resources annotation," in *Computer Science and Information Systems (FedCSIS) 2011 Federated Conference on IEEE*, pp. 973-980, Sept., 2011.
 [10] Eisen, Michael B., Spellman, Brown, and Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences*, Vol.95, No.25, pp.14863-14868, 1998.
 [11] Buyya, Rajkumar, Rajiv Ranjan, and Rodrigo N. Calheiros, "Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services," In *Algorithms and architectures for parallel processing*, Springer Berlin Heidelberg, pp.13-31, 2010.
 [12] Mika, Peter and Giovanni Tummarello, "Web semantics in the clouds," *Intelligent Systems, IEEE*, Vol.23, Issue.5, pp.82-87, 2008.
 [13] Rodriguez-Garcia, Miguel Angel, et al. "Ontology-based annotation and retrieval of services in the cloud," *Knowledge-Based Systems* 56, pp.15-25, Jan., 2014.
 [14] Li Liu, Xiaofen Yao, Liangjuan Qin, and Miao Zhang, "Ontology-based Service Matching in Cloud Computing,"

in *Fuzzy Systems, 2014 IEEE International Conference on IEEE*, pp.2544-2550, Jul., 2014.

[15] Qingtao Wu, Min Cui, Mingchuan Zhang, Ruijuan Zheng, and Ying Lou, "A Cloud Service Resource Classification Strategy Based on Feature Similarity," *Journal of Networks*, Vol.9, No.11, pp.2987-2993, Nov., 2014.
 [16] Openstack [Internet], <http://www.openstack.org/>.
 [17] Sookyoung Park, Hyejeong Kang, Yoonhee Kim, Chongam Kim, and Yunjung Hyun, "An Integrated Scientific Experiment Framework for Numerical Analysis in e-Science Environment," *Computation Tools 2012, Nice, France July 22-27, 2012*.
 [18] Rodrigo N., Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya. "Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," *Software: Practice and Experience*, Vol.41, Issue.1, pp.23-50, 2011.
 [19] W. Cirne, F. Brasileiro, J. Sauvi, Na. Andrade, D. Paranhos, E. Santos-Neto, and R. Medeiros, "Grid Computing for Bag of Jobs Applications," *Proceedings of the 3rd IFIP Conference on E-Commerce, E-Business and E-Government*, Sep., 2003.
 [20] Abdelsamie, Abouelmagd H., and Changhoon Lee, "Decaying versus stationary turbulence in particle-laden isotropic turbulence: Turbulence modulation mechanism," *Physics of Fluids* 24, 2012.



안 윤 선

e-mail : ahnysun@sookmyung.ac.kr
 2013년 숙명여자대학교 컴퓨터과학부(학사)
 2015년 숙명여자대학교 컴퓨터과학부 석사
 관심분야: 클라우드 컴퓨팅, 온톨로지, 지능형 시스템



김 윤 희

e-mail : yulan@sookmyung.ac.kr
 1991년 숙명여자대학교 전산학과(학사)
 1996년 Syracuse University 전산학과 (석사)
 2000년 Syracuse University 전산학과 (박사)
 1991년~1994년 한국전자통신연구소(ETRI) 연구원
 2000년~2001년 Rochester Institute of Technology 컴퓨터공학과 조교수
 2001년~2004년 숙명여자대학교 컴퓨터과학과 조교수
 2004년~2009년 숙명여자대학교 컴퓨터과학과 부교수
 2009년~현 재 숙명여자대학교 컴퓨터과학부 교수
 관심분야: 그리드/클라우드 컴퓨팅 환경, 위크로플로우 제어, 그리드/클라우드 관리