

유전자알고리즘과 신경망을 이용한 웹 서비스 정보

클러스터링

황중연^o 유춘식 김용성
전북대학교 일반대학원 컴퓨터정보학과
{ssing9^o, csyoo, yskim}@chonbuk.ac.kr

Web Service Information Clustering

using Genetic Algorithm and Neural Network

JoongYoun Hwang^o Chun Sik Yoo YongSung Kim.
Computer & Information Science, Chonbuk National University

요 약

오늘날, 웹서비스를 이용한 정보 검색을 하기 위해서는 UDDI 레지스트리의 전문적인 지식이 필요하다. 즉 웹 서비스를 명세하기 위해 사용된 카테고리과 이에 대한 값, 이름 등을 사전에 알고 있어야 한다. 그러나 일반 사용자들은 이러한 사전지식을 충분히 알고 있지 못하면 웹서비스에 대한 정보 검색을 쉽게 할 수 없다. 그러므로 일반 사용자들을 위해 웹서비스에 대한 정보를 카테고리에 맞게 분류하여 검색을 용이하게 할 수가 있다. 따라서 본 논문에서는 보다 효율적으로 웹서비스 정보를 분류하기 위해서 유전자 알고리즘과 신경망을 이용한 클러스터링 기법을 제안하는데 목적이 있다.

핵심어 : 클러스터링, 유전자알고리즘, 신경망, UDDI

1. 서 론

최근 인터넷 관련 기술과 컴퓨터 관련 기술의 발달로 네트워크 상에 분산되어 있는 자원과 정보를 보다 효과적으로 활용하기 위한 방안으로 분산 시스템에 대한 연구가 활발히 진행되고 있다. 이러한 분산 환경에서의 분산응용 프로그램 개발을 지원하기 위해, 최근 W3C에서는 분산 환경에서의 표준 프로그래밍 모델로 웹 서비스(Web Service)라는 개념을 제시하였다[1].

현재 웹 서비스를 이용하여 자원과 정보를 발견하기 위해서는 UDDI 레지스트리에 대한 전문적인 지식이 필요하다. 즉 웹 서비스를 명세하기 위해 사용된 카테고리(비즈니스 카테고리, 서비스 카테고리, tModels 카테고리)와 이에 대한 값, 이름(비즈니스 이름, 서비스 이름, tModels 이름)을 사전에 알고 있어야 한다. 그러나 일반 사용자들의 경우에는 이러한 사전 지식을 충분히 알고 있지 못하므로 키워드를 이용하는 일반적인 정보검색 기법처럼 웹 서비스에 대한 키워드 기반의 검색기법이 필요하다. 즉 서비스 요청자가 필요로 하는 최적의 웹 서비스를 검색하기 위해서는 웹 서비스를 명세하기 위해 사용된 카테고리과 이름, 웹 서비스 명세에 사용된 키워드 등을 함께 사용하는 새로운 방법론이 필요하다. 키워드를 이용한 UDDI 레지스트리 검색 기법에 대한 연구는 국내외적으로 전무한 실정이다.

본 논문의 목적은 사용자로 하여금 효율적인 웹 서비스 정보들을 검색하기 위해서 이 정보들을 유전자알고리즘과 신경망을 이용하여 클러스터링 하는데 있다.

2. 관련연구

유전자알고리즘과 신경망 개념을 이용한 키워드 기반의 지능형 UDDI 레지스트리 검색기법은 웹 서비스 발견을 위한 새로운 방법론이며, 분산응용 개발의 필요성이 증대되고 있는 현대 사회에서 매우 중요한 핵심적 기술 중의 하나이다.

UDDI 레지스트리 검색에 관한 관련연구는 [2]이 있으며, 유전자알고리즘을 이용한 정보검색에 관한 연구는 [3],[4]가 있다. 그리고 신경망을 이용한 정보 검색에 관한 연구로는 [5]이 있다.

[2]는 XML을 기반으로 진보한 UDDI B2B 통합을 위한 방법을 탐색 하였고, [3]는 집단의 크기를 크게 유지하면서 적합도 평가 과정을 줄이는 방안으로 클러스터링에 기반한 효율적인 유전자 알고리즘을 제시하였다.

[4]는 클러스터링 문제를 해결하기 위해서 유전자알고리즘을 사용하여, SICM, STCM, CSPM 3가지 모델을 사용하여 기술하였다. 특히, 이 논문에서는 실험 하고자 하는 p-Median 문제를 풀기 위하여 CSPM을 적용 하였다.

[5]은 동적인 그룹 사용자들에게 적합한 공명이론을 표준으로 한 신경망을 웹에 접근하는 기초로 하였다.

3. UDDI 레지스트리 구조, 유전자알고리즘/신경망을 이용한 클러스터링 기법

본 논문에서 사용된 기반 기술인 UDDI 레지스트리와 유전자알고리즘과 신경망을 이용한 클러스터링 기법에 대하여 기술한다.

3.1 UDDI 레지스트리 구조

UDDI 레지스트리는 서비스 제공자에 대한 정보를 포함하는 businessEntity, 서비스 제공자가 제공하는 웹 서비스들에 대한 businessService, 웹 서비스에 접근하기 위한 정보를 포함하는 bindingTemplate, 웹 서비스를 이용하기 위한 기술적인 정보나 웹 서비스의 분류/식별 체계에 대한 정보인 tModel로 구성된다. UDDI의 주요 데이터 구조를 살펴보면 <그림 1>와 같다.

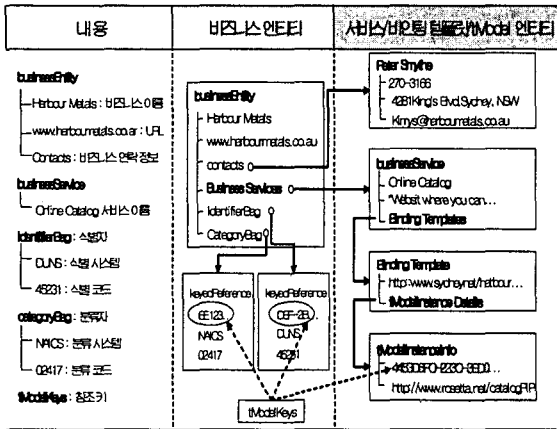


그림 1 UDDI의 데이터 구조

본 논문에서는 UDDI 레지스트리에 저장된 정보로부터 클러스터링을 위한 검색체를 추출하는데 이 UDDI 레지스트리 내의 tModelKey와 description 엘리먼트에서 추출하여 사용한다.

3.2 유전자알고리즘/신경망을 이용한 클러스터링 기법

3.2.1 유전자알고리즘을 이용한 클러스터링 기법

유전자알고리즘을 이용한 클러스터링 기법에 적용시킨 방법을 보면 <그림 2>와 같다.

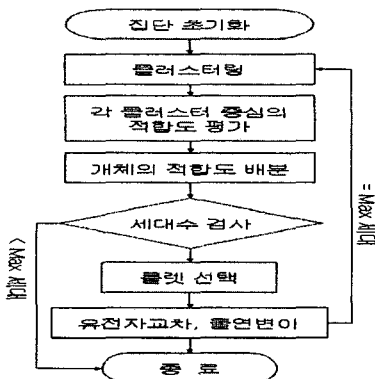


그림 2 클러스터링 유전자 알고리즘

기존의 유전자 알고리즘에서는 개체의 수가 충분하지 않을 때 지역해에 빠지며, 이를 해결하기 위해 개체의 수를 크게 유지해야 하는데 큰 비용이 든다는 문제점을 해결하기 위해서 개체의 유사도를 기준으로 클러스터링을 하고

클러스터링 중심의 적합도만을 평가하고 클러스터 내의 개체들의 유사도에 비례하여 적합도를 제시하며 <그림 2>의 클러스터링 알고리즘을 제안하고 있다.

본 논문에서는 기존의 유전자 알고리즘에 UDDI 레지스트리에서 추출한 키워드들의 가중치를 적용하여 나타난 검색체들을 신경망을 통해 클러스터링 하는데 사용한다.

3.2.2 신경망을 이용한 클러스터링 기법

신경망을 이용한 클러스터링 기법에 적용시킨 방법은 <그림 3>과 같다.

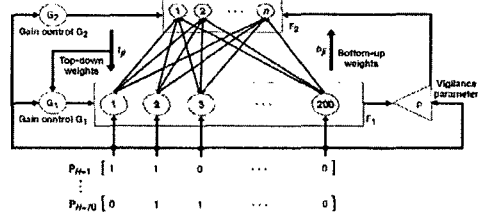


그림 3 ART1 기반의 클러스터링 구조

입력층(F_1)에 1~200개의 URL 벡터가 입력되어 수식에 의한 인식층(F_2)에서 구성된 각각의 고정들이 클러스터를 재구성하고 카테고리에 맞지 않으면 G_1 를 통해 다시 F_1 으로 와서 P 에서 비교해 또다시 F_2 에서 클러스터링 되어지는 방법을 제안하고 있다.

본 논문에서는 클러스터링의 기본구조로 신경망을 사용하고 신경망의 매개변수들을 최적화하기 위해 유전자 알고리즘을 사용 하였다.

4. 유전자알고리즘과 신경망을 이용한 웹 서비스 정보 클러스터링

4장에서 유전자알고리즘과 신경망을 이용하여 웹 서비스 정보를 클러스터링하는 기법을 제안한다.

4.1 검색체의 구성

UDDI 레지스트리에서 추출된 엘리먼트를 대상으로 <그림 4>와 같이 검색체를 구성하여, 웹 서비스 정보를 클러스터링하기 위한 초기 작업을 한다.

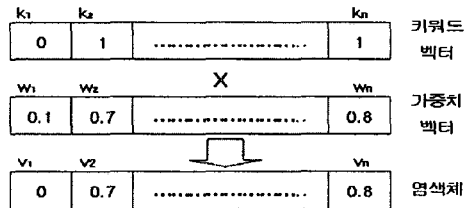


그림 4 검색체 구성 예

UDDI 레지스트리에서 추출한 각각의 키워드를 k_1, k_2, \dots, k_n 으로 하고 이 추출한 각각의 키워드들에 대해 중요도 및 유사도의 가중치를 산출한 가중치 벡터와 키워드 벡터를 곱해 각각의 검색체 v_1, v_2, \dots, v_n 을 구성 하였다.

4.2 신경망의 구성

UDDI 레지스트리에서 추출된 영색체를 신경망에 이용하여 웹 서비스의 정보들의 클러스터를 결정하는 신경망의 구조는 <그림 5>와 같다.

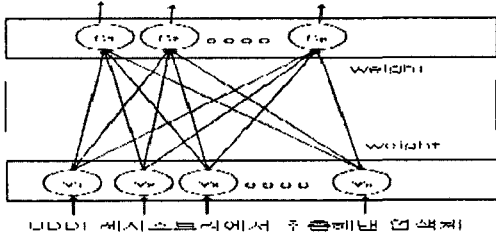


그림 5 신경망의 예

UDDI 레지스트리에서 추출된 영색체를 신경망에 이용하기 위해 입력층에 추출된 영색체 v_1, v_2, \dots, v_n 를 갖으며, 이러한 입력 값이 들어 왔을 때 초기신경망을 구성하고 유전자 알고리즘을 통해 클러스터에 맞는 영색체를 추출하며 신경망을 재구성함으로써 나타나는 정보로 클러스터링을 수행하여 웹 서비스 정보들의 클러스터 c_0, c_1, \dots, c_p 를 결정하였다.

4.3 클러스터링 알고리즘

<그림 6>은 본 논문에서 제안하고 있는 유전자알고리즘과 신경망의 개념을 이용하여 웹 서비스 관련 정보를 클러스터링 하기 위한 개괄적인 알고리즘이다. 즉 지능형 UDDI 레지스트리 검색을 위해 유전자 알고리즘과 신경망을 이용하여 웹 서비스 정보에 대한 클러스터링을 수행하며, 이때 클러스터링의 대상이 되는 정보는 확장된 UDDI 데이터 구조에서 추출한 웹 서비스에 대한 논리적인 정보이다.

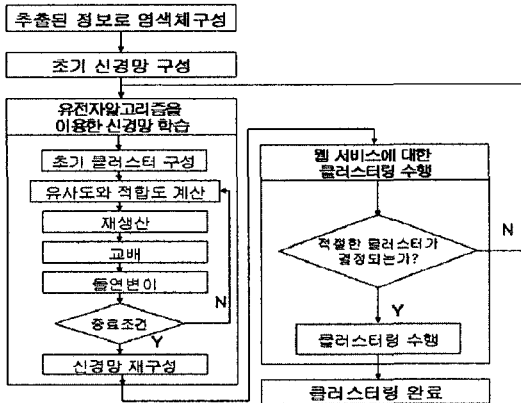


그림 6 유전자알고리즘과 신경망을 이용한 정보 클러스터링 알고리즘

4.4 웹 서비스 검색

UDDI 레지스트리에 있는 정보를 검색하기 위해 BusinessEntity와 BusinessService 그리고 tModel 정보를 이용하여 WSDL 문서를 작성한 후 정해진 UDDI 레지스트리에 메시지를 보낸 후 검색된 정보를 대상으로 영색체와 신경망 구조

를 적용하여 로컬 시스템에 UDDI 정보 검색을 구현하는 개략적인 시스템 구조도는 <그림 7>과 같다.

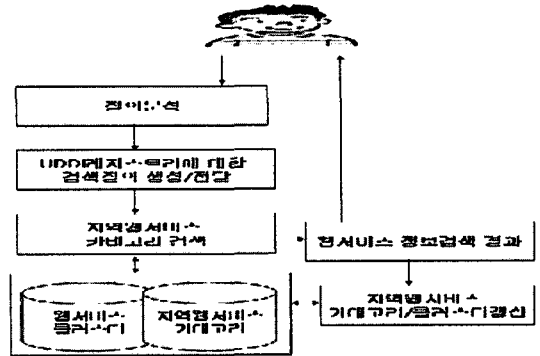


그림 7 웹서비스 검색 시스템 구조도

5. 결론 및 향후 연구과제

본 논문에서 연구된 내용은 다음과 같이 나누어 진다. 첫째, 유전자 알고리즘을 이용한 웹 서비스 UDDI 레지스트리의 영색체 추출/구성과 둘째, 추출된 영색체를 신경망에 적용하여 웹 서비스를 클러스터링 하는 방법을 제안하였다.

클러스터링을 할 때 유전자 알고리즘과 신경망을 이용함으로써 사용자 하여금 웹 서비스 정보에 대한 효율적으로 카테고리별 검색을 할 수 있을 뿐만 아니라 더욱 다양하고 많은 자료를 검색 수행할 수 있을 것이다.

또한 유전자알고리즘과 신경망을 이용한 UDDI 레지스트리의 정보를 내용별로 클러스터링하여 로컬 DB를 구축함으로써 UDDI 정보 검색자 들에게 편의를 제공하고 웹 서버의 기능을 활성화하여 기업간의 활발한 전자상거래를 향유로 높이기 위해서 본 논문을 제안하였다.

향후 연구과제로는 본 논문에서 제안하고 있는 유전자알고리즘과 신경망을 이용하여 웹 서비스 정보를 클러스터링의 알고리즘 상세화, 구현, 실험/평가를 통한 효율성일 입증하겠다.

참고문헌

[1] Ying Huang, Jen-Yao Chung., "A Web services-based, framework for business integration solutions," Electronic Commerce Research and Applications, Vol.2, 2003.
 [2] LIANG-JIE ZHANG, TIAN CHAO, JEN-YAO CHUNG., "XML-Based Advanced UDDI Search Mechanism for B2B Integration." Electronic Commerce Research, Vol.3, pp.25-42, 2003.
 [3] 원홍희, 조성배, "클러스터링 기반의 효율적 유전자 알고리즘의 체계적인 성능 평가", 정보과학회 학술발표논문집, 제29권, 제1호, pp.298-300, 2002.
 [4] Yu-Chiun Chiou, Lawrence W.LAN., "Genetic clustering algorithms," European Journal of Operational Research, Vol.135, pp.413-427, 2001.
 [5] Santosh k., "Adaptive Neural Network Clustering of Web Users," IEEE, Vol.4, pp.34-40,2004.