

지식모니터링시스템에서 감성기준을 고려한 EFASIT 모델

An EFASIT model considering the emotion criteria in Knowledge Monitoring System

류 경 현* 피 수 영**
Kyung-Hyun Ryu Su-Young Pi

요 약

웹의 등장은 전통적인 정보검색을 비롯하여 지식관리와 일반 상거래 등 사회 전 분야의 급격한 변혁을 초래하였다. 그러나 검색엔진은 일반적으로 관련된 계산함수에 의해 순서화된 URL의 방대한 목록을 제공하지만, 관련 없는 정보의 필터링이나 사용자가 필요로 하는 정보의 검색에 많은 시간이 소요된다.

본 논문에서는 웹상의 효율적인 문서검색을 위해서 영역 코퍼스 정보를 바탕으로 확장된 퍼지 계층화 의사결정법(Extended Fuzzy AHP Method : EFAM)과 유사도 기법(Similarity Technology : SIT)을 결합하고, 감성기준을 고려한 EFASIT(Extended Fuzzy AHP and Similarity Technology)모델을 제안한다.

제안한 감성기준을 고려한 EFASIT 모델은 다양한 의사결정자들의 퍼지 지식의 통합으로 좀 더 명확한 규칙을 생성할 수 있고 의사결정을 하는데 도움을 준다는 것을 실험을 통하여 확인한다.

ABSTRACT

The appearance of Web has brought a substantial revolution to all fields of society such knowledge management and business transaction as well as traditional information retrieval.

In this paper, we propose an EFASIT(Extended Fuzzy AHP and Similarity Technology) model considering the emotion analysis. And we combine the Extended Fuzzy AHP Method(EFAM) with Similarity Technology(SIT) based on the domain corpus information in order to efficiently retrieve the document on the Web. The proposed the EFASIT model can generate the more definite rule according to integration of fuzzy knowledge of various decision-maker, and can give a help to decision-making, and confirms through the experiment.

☞ keyword : Fuzzy Analytic hierarchy process(퍼지계층화 의사결정법), similarity relation(유사관계), data fusion(데이터결합), Knowledge Monitoring system(지식 모니터링 시스템)

1. 서 론

의사결정의 전 과정을 여러 단계로 나눈 후 이를 단계별로 분석 해결함으로써 합리적인 의사결정에 이를 수 있도록 지원해주는 방법으로 AHP(Analytic Hierarchy Process)가 있다. 그러나 AHP는 평가치로 크리스프 값(crisp value)을 사용함에 따라 평가대안에 대한 평가자의 판단이 정

확하게 정량화되지 못하고 평가에 있어 주로 분명한 의사결정문제에만 사용된다. 평가자의 판단을 수치로 정확히 변환하는데 있어 불확실한 요소들을 고려하지 않고 있으며 평가에 있어서도 균등하지 못한 크기로 평가가 될 수 있다[1].

이를 보완한 것이 Fuzzy AHP로 정확한 크리스프 값 대신에 퍼지 언어변수들을 삼각 퍼지숫자에 대응시킴으로써 매우 주관적인 판단에 대해 비수치적 언어변수를 사용하는 전문가의 선호와 관련된 부정확성을 통합할 수 있다[2]. 그러나 데이터의 관점에서 이 방법은 퍼지 주관적 속성 가중치만을 고려함으로써 계산의 복잡성을 가지고

* 정 회 원 : 대구대학교 시간강사
ryu0919@naver.com

** 정 회 원 : 대구가톨릭대학교 실용컴퓨터 책임교수
sypi@cu.ac.kr

[2011/01/26 투고 - 2011/02/21 심사 - 2011/05/24 심사완료]

있다. 의사결정 과정에서 기존의 방법들은 복잡하고 다수 전문가들의 가중치를 적용하지 않아서 전문가들의 능력에 따른 가중치를 부여하여 할 의사결정에 있어서는 정확한 의사결정을 하는데 문제가 나타날 수 있다.

따라서 본 논문에서는 의사결정에 관한 요소를 계층화 구조로 나타내는 기존의 방법들을 데이터의 관점에서 대안, 평가기준, 주관적 속성가중치, 개념과 객체 사이의 퍼지 관계를 기반으로 하는 EFASIT(Extended Fuzzy AHP and Similarity Technology)모델을 제안한다.

2. 관련 연구

2.1 Fuzzy AHP

Fuzzy AHP는 기본적으로 기존 AHP와 같지만 연산과정에 사용되는 데이터가 크리스프 값이 아닌 퍼지 숫자라는 점이 다르다. 삼각퍼지 숫자는 세 개의 점으로 표현할 수 있기 때문에 사용이 간편하다.

Fuzzy AHP의 과정은 다음과 같이 나타낸다.

단계 1 : 계층구조 구성

단계 2 : 퍼지 판단행렬 \bar{V} 구성

퍼지 판단행렬 \bar{V} 는 각 대안과 평가기준사이의 쌍대비교행렬이다. 두 기준 중의 어떤 것이 더 중요한가를 요구함으로써 쌍대비교를 위해 언어 변수들을 식 (1)에 배정한다.

$$\bar{V} = \begin{pmatrix} \hat{1} & \hat{v}_{12} & \cdots & \hat{v}_{1n} \\ \hat{v}_{21} & \hat{1} & \cdots & \hat{v}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{v}_{n1} & \hat{v}_{n2} & \cdots & \hat{1} \end{pmatrix} = \begin{pmatrix} \hat{1} & \hat{v}_{12} & \cdots & \hat{v}_{1n} \\ \hat{v}_{21}^{-1} & \hat{1} & \cdots & \hat{v}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{v}_{n1}^{-1} & \hat{v}_{n2}^{-1} & \cdots & \hat{1} \end{pmatrix} \quad (1)$$

여기서

$$\hat{v}_{ij} = \begin{cases} 1, 3, 5, 7, 9 & : \text{기준 } i \text{는 기준 } j \text{에 대해 상대적으로 중요} \\ \hat{1} & : i=j \\ \hat{1}^{-1}, \hat{3}^{-1}, \hat{5}^{-1}, \hat{7}^{-1}, \hat{9}^{-1} & : \text{기준 } i \text{는 기준 } j \text{에 대해 상대적으로 작게 중요} \end{cases}$$

단계 3 : 각 기준의 퍼지 가중치를 계산

각 기준의 퍼지 가중치는 식 (2), 식 (3)과 같이 계산한다.

$$\hat{g}_i = [\hat{v}_{i1} \otimes \hat{v}_{i2} \otimes \cdots \otimes \hat{v}_{in}]^{\frac{1}{n}}, \forall i=1,2,\dots,n \quad (2)$$

$$\hat{w}_i = \frac{\hat{g}_i}{\hat{g}_1 \oplus \hat{g}_2 \oplus \cdots \oplus \hat{g}_n} \quad (3)$$

여기서 \hat{v}_{ij} 는 기준 j 에 대한 기준 i 의 퍼지 쌍대값이다. \hat{g}_i 는 각 기준에 대한 기준 i 의 퍼지 쌍대값의 기하평균이고 \hat{w}_i 는 i 번째 기준의 퍼지 가중치이다.

단계 4 : 계층 레이어 순서

각 대안의 최종 퍼지 가중치는 식 (4)와 같이 계층 레이어 순서에 의해 계산된다.

$$\bar{U}_i = \sum_{j=1}^n \hat{w}_j \cdot \hat{g}_{ij} \quad (4)$$

여기서 \hat{g}_{ij} 는 i 번째 대안에 대한 j 번째의 퍼지 가중치이다. \bar{U}_i 는 삼각 퍼지숫자 $\bar{U}_i=(l, m, u)$ 에 의해 나타낼 수 있다.

단계 5 : 대안 순위

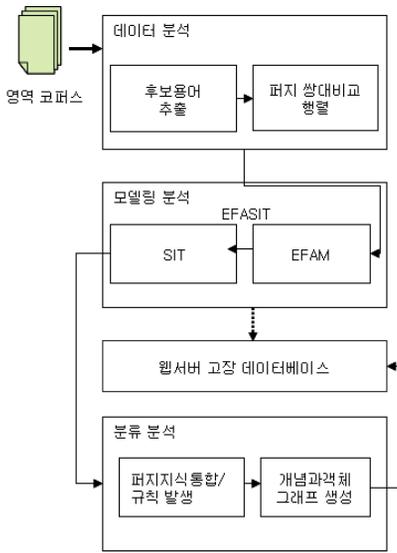
대안들의 최종 퍼지 값들은 퍼지 숫자에 의해 표현된다. 최적 대안을 선택하기 위하여 퍼지 숫자로부터 크리스프 값에 대한 방법을 정의하는 것이 필요하다.

3. EFASIT 모델

본 논문에서는 ‘웹 서버에서 발생할 수 있는 고장’ 문제 영역에 대하여 온톨로지 개념에 기반한 EFAM과 SIT를 결합한 감성분석 결과를 도입한 EFASIT 모델을 제안한다.

3.1 EFASIT 모델의 구성

온톨로지 개념에 기반한 EFASIT 모델은 (그림 1)과 같이 프로세스에 대한 합리적인 프레임워크 내에 추론을 하기 위한 구성요소들의 집합을 가진 이론적인 패러다임으로 주요한 개념들과 상호관계를 포함하는 체계를 제공하며 데이터 분석, 모델링 분석 및 분류 분석의 세 가지 구성요소를 가진다.



(그림 1) 온톨로지 개념에 기반한 EFASIT 모델

첫 번째 데이터 분석은 웹상에서 효율적인 문서검색을 위하여 영역 코퍼스에서 이용할 수 있는 문서들로부터 후보 개념들을 추출(특정한 웹사이트 또는 가상 커뮤니티에서 교환된 문서들)하여 퍼지 쌍대비교행렬을 사용하여 개념들 간의 관계를 나타낸다.

두 번째 모델링 분석은 추출된 개념들 중에서 EFAM과 SIT를 구하여 웹서버 고장 데이터베이스에 저장한다. EFAM은 다수 전문가의 가중치를 정의하기 위한 관련 기준 중 하나인 감성기준을 포함하여 개념들의 최종 개념 가중치를 계산한다. SIT는 전문가별 통합된 개념 유사도와 소속도로 구분하여 개념들 간의 근접관계를 변환에 의한

유사도, min 연산자, 다이스 계수, 코사인 유사도, 확장 자카드 계수를 사용하여 계산한다.

세 번째 분류 분석은 EFAM과 각 SIT를 결합하여 개념들을 분류하고 다수 전문가들에 의해 구해진 개념과 객체간의 퍼지지식을 통합하고 확신도와 퍼지값을 구하여 규칙을 생성한다.

3.2 EFASIT모델의 모델링

의사결정 과정에서 기존의 방법들은 다수 전문가들의 가중치를 적용하지 않았으며 만일 전문가들의 능력에 따른 가중치를 부여할 경우에 이를 반영하여 처리할 방법의 부재를 문제점으로 지적할 수 있다. 즉 전문가들의 의견의 불일치에 대한 문제를 해결하기 위해 여기서는 EFAM을 제안한다.

(1) EFAM의 최종 개념가중치 계산과 SIT

다수 전문가는 다른 과거경험, 배경지식, 관련 평판 및 감성을 가지고 있기 때문에 관련 기준을 기반으로 전문가별 가중치를 정의하는 것이 필요하다. 여기서 ‘감성’ 기준은 ‘웹 서버에서 발생할 수 있는 고장’의 문제가 발생했을 때 자신의 진정한 기분을 자각하고 이를 진심으로 납득할 수 있는 결단을 내릴 수 있는 능력을 말한다[3,4] 이 단계에서 전문가들은 앞에서 제시된 기준들 중 감성기준 고려여부에 의해 자기-특점을 기록한다. 그리고 전문가별 가중치는 식 (5)와 식 (6)을 이용함으로써 계산된다.

$$[e_1, e_2, \dots, e_n] = [c_1, c_2, \dots, c_m] \cdot \begin{bmatrix} sc_{11} & \dots & sc_{1j} & \dots & sc_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ sc_{i1} & \dots & sc_{ij} & \dots & sc_{in} \\ \dots & \dots & \dots & \dots & \dots \\ sc_{m1} & \dots & sc_{mj} & \dots & sc_{mn} \end{bmatrix} \quad (5)$$

$$[\hat{e}_1, \hat{e}_2, \dots, \hat{e}_j, \dots, \hat{e}_n] = \left[\frac{e_1}{E}, \frac{e_2}{E}, \dots, \frac{e_j}{E}, \dots, \frac{e_n}{E} \right] \quad (6)$$

여기서 $E = e_1 + e_2 + \dots + e_n$

식 (5)에서 c_i 는 각 기준의 가중치이고 sc_{ij} 는 i 번째 기준에 대한 j 번째 전문가에 의해 주어진 자기-특점이다. 그리고 \hat{e}_j 는 식 (6)에서 모든 전문가들(즉, E)의 합에 관하여 e_j 에 대한 비를 나타낸다. 다수 전문가에 의해 주어진 각 개념들의 상대 중요도에 감성기준을 고려한 전문가들의 최종 개념가중치는 식 (7)에 나타낸다.

$$[fw_1, \dots, fw_k, \dots, fw_m] = [e_1, e_2, \dots, e_j, \dots, e_n] \cdot \begin{bmatrix} \hat{w}_{11} & \dots & \hat{w}_{1k} & \dots & \hat{w}_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{w}_{j1} & \dots & \hat{w}_{jk} & \dots & \hat{w}_{jm} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{w}_{n1} & \dots & \hat{w}_{nj} & \dots & \hat{w}_{nm} \end{bmatrix} \quad (7)$$

여기서 $\hat{w}_{jk}, j=1,2,\dots,n, k=1,2,\dots,m$ 에 대해, j 번째 전문가에 의해 주어진 k 번째 개념의 평균가중치를 나타낸다. 그리고 fw_k 는 행렬을 곱한 후 개념의 최종 가중치를 표현한다.

개념들 간의 근접관계를 구하기 위하여 통합된 개념 유사도와 소속도로 구분하여 변환에 의한 유사도, min 연산자, 다이스 계수, 코사인 유사도, 확장 자카드 계수를 사용하여 비교한다 [5-8].

3.3 EFASIT 모델의 개념 분류분석

데이터 결합(data fusion)은 여러 다른 SIT들을 결합하거나 다른 유형의 색인어들을 결합함으로써 검색 성능을 높이려는 전략이다. 데이터 결합은 단일 검색 모형이나 단일 유형의 색인어를 이용하여 얻을 수 있는 검색 성능의 한계를 뛰어넘고자 하는 시도이다[9,10]. 개념 분류 분석에는 EFAM에서 계산한 최종 개념 가중치(fw_i)와 변환에 의한 SIT, min 연산자 SIT, 다이스 계수 SIT, 코사인 유사도, 확장 자카드 계수에서 구한 각각의 유사도들과 결합한 다음 식 (8)에 의하여 개념 i 에 대하여 가중치된 개념 값 즉 분류값 cv_i 를 계산한다.

$$cv_i = \sum_{j=1}^n fw_j \cdot s_{ij}^c \quad (8)$$

여기서 n 은 분석에서 사용되는 기준의 수이고 fw_j 는 기준 j 에 대해 찾은 최종 중요도이다. EFASIT는 cv_i 값을 고려하여 개념들을 분류한다.

3.3.1 개념과 객체 그리드 평가

다수 전문가들로부터 공통개념그리드를 도출한 후 지식 통합을 하기 위해서 개념과 객체그리드를 평가한다. 여기서 객체(object)란 검색할 문제, 도달할 목표, 해결할 의사결정 또는 실행할 행동 등을 나타낸다.

(표 1) 개념과 객체 그리드의 예

개념이름	객체1	객체2	...	객체 n
개념 1	cf_{11}/fv_{11}	cf_{12}/fv_{12}	...	cf_{1n}/fv_{1n}
⋮	⋮	⋮	⋮	⋮
개념 n	cf_{n1}/fv_{n1}	cf_{n2}/fv_{n2}	...	cf_{nn}/fv_{nn}

(표 1)에 나타낸 것과 같이 ‘개념이름’ 열은 공통개념그리드와 동일한데에서 개념과 객체 그리드를 만든다. 전문가들은 개념 i 와 객체 j 사이의 관계를 설명하기 위하여 개념과 객체 그리드의 각 엔트리에 대한 확신도/퍼지값(즉, $(cf_{ij})/(fv_{ij})$) 쌍을 준다. 영역 개념으로부터 도출된 퍼지평가에 의하여 퍼지지식을 통합하고 규칙을 생성한다.

3.3.2 퍼지지식 통합과 규칙 생성

이 단계에서는 감성기준을 고려한 영역 전문가들에 의해 주어진 퍼지 평가와 값들을 통합하여 규칙을 생성한다. 이것은 식 (9), 식 (10), 식 (11)에 나타낸다.

$$\begin{aligned}
 & [dcf_{1/dfv_1}, dcf_{2/dfv_2}, \dots, dcf_{k/dfv_k}, \dots, dcf_{m/dfv_m}] \\
 & = [\hat{e}_1, \hat{e}_2, \dots, \hat{e}_j, \dots, \hat{e}_n] \\
 & \cdot \begin{bmatrix} dcf_{11}/dfv_{11} & \dots & dcf_{1k}/dfv_{1k} & \dots & dcf_{1m}/dfv_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ dcf_{j1}/dfv_{j1} & \dots & dcf_{jk}/dfv_{jk} & \dots & dcf_{jm}/dfv_{jm} \\ \dots & \dots & \dots & \dots & \dots \\ dcf_{n1}/dfv_{n1} & \dots & dcf_{nk}/dfv_{nk} & \dots & dcf_{nm}/dfv_{nm} \end{bmatrix} \quad (9)
 \end{aligned}$$

$$\begin{aligned}
 cf &= [dcf_1, dcf_2, \dots, dcf_k, \dots, dcf_m] \\
 & \cdot [fw_1, \dots, fw_k, \dots, fw_m] \quad (10)
 \end{aligned}$$

$$fv = round [dfv_1, dfv_2, \dots, dfv_k, \dots, dfv_m] \quad (11)$$

여기서 dcf_{jk}/dfv_{jk} 는 가능한 의사결정에 대하여 j 번째 전문가에 의해 주어진 k 번째 개념의 확신도와 퍼지값을 나타내고 dcf_k/dfv_k 는 가능한 의사결정에 대하여 통합된 확신도와 퍼지값을 나타낸다. cf 는 규칙의 확신도를 나타내고 fv 는 각 dfv_k 를 가장 가까운 정수로 조정하면서 최종 퍼지값을 표현한다.

4. 실험 및 평가

본 논문에서는 온톨로지를 “어휘들에 대해서 일정 영역의 개념적 예들을 한곳으로 집합시킨 하나의 독립된 집합체”로 정의하고 정보는 웹 애플리케이션에서 고장의 원인과 보급에 대하여 cnet.com[11]과 eweek.com[12]과 같은 technology 웹 사이트 상에 리스트 된 웹 사이트 정전 사건과 시스템 고장의 케이스연구를 조사하고 수집하여 고장원인에 대하여 시뮬레이션을 하였다.

그 중에서 ‘웹 서버에서 발생할 수 있는 고장 (Possible failure of Web Server)’의 문제 영역을 고려해 보면 ‘컴퓨터 바이러스’, ‘디스크 공간’, ‘메모리 공간’과 같은 서버고장에 관한 여러 가지 관련된 고장원인들이 있다. 다른 과거 경험, 배경 지식, 관련 평판, 감성을 가진 전문가들은 주로 개인의 서버 고장원인을 정의하기 위하여 익숙한 원인들을 사용하였다.

4.1 EFAM

4.1.1 전문가별 개념 구성

(표 2), (표 3) 및 (표 4)는 전문가 e_1, e_2, e_3 에 의해 제공한 개념 그리드를 나타낸다.

(표 2) 전문가 e1의 개념 그리드

개념이름	단위	언어형태	소속함수	개념설명
하드디스크 공간	M Bytes	부족/보통/충분	(10, 30, 50)	자유 디스크 공간
메모리 공간	M Bytes	부족/보통/충분	(100, 300, 500)	자유 메모리 공간
컴퓨터 바이러스	위험순위	1/2/3/4/5	(1, 3, 5)	Trend micro의 바이러스 정보
하드웨어 손상	논리	참/거짓	N/A	컴퓨터 하드웨어상태
정전	-	낮음/보통/높음	(10, 30, 50)	UPS(Uninterrupted Power Supply) 잔여 파워
프로그램 버그	논리	참/거짓	N/A	프로그램의 동작조건
운영체제 고장	논리	참/거짓	N/A	서버 동작상태
CPU 부하	-	낮음/보통/높음	(80, 90, 100)	CPU 부하

(표 3) 전문가 e2의 개념 그리드

개념이름	단위	언어형태	소속함수	개념설명
네트워크 부하	-	낮음/보통/높음	(50, 70, 100)	네트워크 부하
CPU 부하	-	낮음/보통/높음	(70, 80, 90)	CPU 부하
메모리 공간	M Bytes	부족/보통/충분	(100, 300, 500)	자유 메모리 공간
해커	위험순위	낮음/보통/높음	(1, 3, 5)	서버에 대한 위험 순위
하드웨어 고장	논리	참/거짓	N/A	컴퓨터 하드웨어상태

(표 4) 전문가 e3의 개념 그리드

개념이름	단위	언어형태	소속함수	개념설명
메모리 공간	M Bytes	부족/보통/충분	(100, 300, 500)	자유 메모리 공간
네트워크 부하	-	낮음/보통/높음	(60, 80, 100)	네트워크 부하
정전	-	낮음/보통/높음	(20, 40, 60)	UPS 잔여 파워
프로그램 버그	논리	참/거짓	N/A	프로그램의 동작조건
운영체제 고장	논리	참/거짓	N/A	서버 동작상태
CPU 부하	-	낮음/보통/높음	(60, 70, 80)	CPU 부하

여기서 언어형태 낮음(또는 짧음, 작음, 느림, 부족, 불충분), 보통(또는 평균, 중간) 그리고 높음(또는 김, 큼, 빠름, 충분)은 삼각 퍼지수의 소속 함수로 나타내었다.

4.1.2 전문가별 가중치 계산

각 전문가 e_1, e_2, e_3 는 기존의 세 가지 기준 과거경험, 배경지식, 관련평판에 감성기준을 추가하여 전문가별 가중치를 계산하였다. 설문 결과 과거경험, 배경지식, 관련평판 그리고 감성의 중요한 비는 각각 0.24, 0.27, 0.24, 0.25로 계산되었고 기준의 중요도는 배경지식 > 감성 > 과거경험, 관련평판 순으로 나타났다.

$$[e_1, e_2, e_3] = [0.24, 0.27, 0.24, 0.25] \cdot \begin{bmatrix} 0.8 & 0.5 & 0.9 \\ 0.9 & 0.9 & 0.8 \\ 0.7 & 0.7 & 0.8 \\ 0.6 & 1.0 & 0.7 \end{bmatrix}$$

$$= [0.75, 0.78, 0.8]$$

$$[\hat{e}_1, \hat{e}_2, \hat{e}_3] = [0.32, 0.33, 0.35]$$

전문가 e_1 의 가중치는 0.32이고 전문가 e_2 의 가중치는 0.33, 전문가 e_3 의 가중치는 0.35로 전문가 $e_3 > 전문가 e_2 > 전문가 e_1$ 순으로 나타났다. 그리고 감성기준을 고려하지 않았을 경우, 전문가 e_1 의 가중치는 0.34이고 전문가 e_2 의 가중치는 0.3, 전문가 e_3 의 가중치는 0.36으로 전문가 $e_3 > 전문가 e_1 > 전문가 e_2$ 순으로 나타났다.

EFAM/Fuzzy AHP 및 각 전문가들의 개념 중요도를 분석해 본 결과 가중치가 가장 높았던 전문가 e_3 와 EFAM/Fuzzy AHP의 개념 중요도의 순위 차를 살펴보면 ‘정전’이 4로 가장 큰 차이가 났고 그 다음 ‘프로그램 버그’가 3차이가 나고 ‘메모리 공간’, ‘하드웨어 손상’, ‘CPU 부하’가 2차이 나고 ‘하드디스크 공간’, ‘운영체제 고장’, ‘네트워크 부하’가 1차이가 남을 알 수 있었다. 그리고 Fuzzy AHP는 식 (4)를 사용하여 개념 중요도를 계산하여 분석한 결과 EFAM과는 전문가 e_1 에서

‘메모리 공간’과 ‘컴퓨터 바이러스’의 순위가 다른 것을 알 수 있었다. 따라서 ‘웹서버에서 발생할 수 있는 고장’ 문제 영역의 개념의 중요도는 EFAM과 Fuzzy AHP에서 ‘컴퓨터 바이러스’가 가장 높다는 것을 알 수 있었다.

4.2 SIT

4.2.1 SIT에 의한 개념 분류 비교

(표 5)는 다수 전문가의 통합된 개념 유사도를 사용한 경우인 변환에 의한 유사도에 대해 나타내었다.

(표 5) 전문가별 통합된 개념 유사도

방법	특징
변환에 의한 유사도	<ul style="list-style-type: none"> 원래의 근접도 값이 $[0, \infty]$의 값을 가진 경우에는 비선형 변환이 필요함 새로운 척도에서는 원래 값들이 가졌던 서로간의 관계가 달라지게 됨 예) 상관관계는 구간 $[-1, 1]$ 사이의 값을 가지는 유사도의 척도- 절대값을 이용해 구간 $[0, 1]$로 사상하면 일부 응용에서는 중요할 수도 있는 부호 정보를 잃어버리게 됨
분류	A : 2 / 20 % , B : 3 / 50 % , C : 5 / 100 %

(표 6)은 전문가별 통합된 개념 소속도를 이용하여 각 SIT에 의한 개념 분류를 분석하였다.

전문가별 통합된 개념 소속도를 사용한 경우에 코사인 유사도에 의한 개념 분류가 가장 우수하다는 것을 알 수 있었다.

4.3 EFASIT 모델과 기존 방법

본 논문이 제안한 EFASIT 모델은 ‘웹서버에서 발생할 수 있는 고장’ 문제 영역의 정보에서 EFAM을 사용하여 다수 전문가들의 가중치를 통합하고 개념들의 최종 중요도를 계산하고, SIT들과 결합하여 개념들의 순위 비교에 의해 각 분류

(표 6) 전문가별 통합된 개념 소속도

방 법	특 징
min 연산자	<ul style="list-style-type: none"> 퍼지집합의 사용은 min, max 가중치를 선택하기 때문에 색인어에 가중치를 부여해야만함 검색된 문서의 순위 부여 능력이 모든 검색어에 민감하지 못함
분류	A : 4 / 40 % , B : 3 / 70 % , C : 3 / 100 %
다이스 계수	<ul style="list-style-type: none"> 문서분류의 정확도를 어느 정도 보장 절대값을 이용해 구간 [0,1]로 사상하면 일부 응용에서는 중요할 수도 있는 부호 정보를 잃어버리게 됨 미리 규칙을 위한 학습과정이 필요하며 그에 따른 학습 데이터가 반드시 필요
분류	A : 5 / 50% , B : 4 / 90% , C : 1 / 100%
코사인 유사도	<ul style="list-style-type: none"> 객체들간의 다수의 코사인 유사도를 계산할 경우에 객체들을 단위길이로 정규화하면 계산시간이 감소
분류	A : 6 / 60% , B : 3 / 90% , C : 1 / 100%
확장 자카드 계수	<ul style="list-style-type: none"> 문서분류의 정확도를 어느 정도 보장 미리 규칙을 위한 학습과정이 필요하며 그에 따른 학습 데이터가 반드시 필요
분류	A : 1 / 10% , B : 1 / 20% , C : 3 / 50% , D : 4 / 90% , E : 1 / 100%

(표 7) EFASIT 모델과 기존 방법의 비교

단계	AHP	Fuzzy AHP	본 논문의 EFASIT 모델(EFAM+SIT)
1	의사결정 계층 설정	계층구조 구성	다수 전문가로부터 개념 구성
			공통 개념 그리드 형성
2	의사결정 요소들간의 쌍대비교(크리스프 값)	퍼지 판단행렬 구성 (퍼지숫자)	전문가별 가중치 계산(기하평균/가중치)
			퍼지 쌍대비교 행렬 구성(퍼지숫자)
3	의사결정 요소들의 상대적 가중치 유도 (고유값/가중치)	각 기준의 퍼지 가중치를 계산 (기하평균/가중치)	일관성 비율 분석
			각 개념의 상대적 중요도 계산
4	판단의 일관성 측정	계층 레이어 순서	각 개념의 최종 가중치 계산(감성기준 고려여부에 따라 전문가별 가중치를 통합)
			민감도 분석
5	계층구조의 종합화를 통한 대안들의 우선순위 및 가중치 결정	대안순위	SIT (통합된 개념유사도/소속도에 의한 방법)
			EFAM과 SIT의 결합
6(확장단계) 개념과 객체 그리드의 지식 통합표현	없음	없음	개념과 객체 그리드 평가(확신도/퍼지값)
			퍼지지식 통합과 규칙생성

방법들을 분석한 결과 EFAM과 min 연산자 분류 방법이 문서추출에 있어서 비교적 좋았다. 그리고 각 전문가들의 개념과 객체 그리드를 평가하여 퍼지지식을 통합해서 새로운 규칙을 생성함

으로써 웹 자원을 효과적으로 관리하고 더 명확한 의사결정을 할 수 있도록 한다. (표 7)은 EFASIT 모델과 기존 방법을 단계별로 비교하여 나타내었다.

(표 8) EFAM/Fuzzy AHP와 SIT를 결합한 순위 비교

	하드디스크 공간	메모리 공간	컴퓨터 바이러스	하드웨어 손상	정전	프로그램 버그	운영체제 고장	CPU 부하	네트워크 부하	해커
EFAM/Fuzzy AHP 최종 중요도	0.127	0.141	0.222	0.080	0.061	0.114	0.061	0.063	0.065	0.068
EFAM/Fuzzy AHP 순위	3	2	1	5	9	4	10	8	7	6
분류값1	0.406	0.408	0.828	0.260	0.158	0.208	0.167	0.205	0.234	0.113
분류값1의 순위	3	2	1	4	9	6	8	7	5	10
EFAM/Fuzzy AHP 순위차	-	-	-	1	-	-2	2	1	2	-4
분류값2 (감성기준 고려)	0.478	0.507	0.776	0.297	0.182	0.424	0.230	0.217	0.183	0.190
분류값2의 순위	3	2	1	5	10	4	6	7	9	8
EFAM/Fuzzy AHP와 순위차	-	-	-	-	-1	-	4	-1	-2	2
분류값2 (감성기준 고려안함)	0.474	0.513	0.776	0.299	0.182	0.428	0.229	0.214	0.180	0.190
분류값2의 순위	3	2	1	5	9	4	6	7	10	8
EFAM/Fuzzy AHP 순위차	-	-	-	-	-	-	4	1	-3	-2
분류값3	0.215	0.185	0.478	0.157	-0.189	0.193	0.065	0.151	0.040	0.011
분류값3의 순위	2	4	1	5	10	3	7	6	8	9
EFAM/Fuzzy AHP 순위차	1	-2	-	-	-1	1	3	2	-1	-3
분류값4	1.116	1.225	1.951	0.692	0.430	1.092	0.507	0.558	0.516	0.578
분류값4의 순위	3	2	1	5	10	4	9	7	8	6
EFAM/Fuzzy AHP 순위차	-	-	-	-	-1	-	-1	-1	-1	1
분류값5	1.002	1.155	1.884	0.660	0.344	1.103	0.470	0.520	0.445	0.551
분류값5의 순위	4	2	1	5	10	3	8	7	9	6
EFAM/Fuzzy AHP 순위차	-1	-	-	-	-1	1	2	-1	-2	-

4.4 EFAM/Fuzzy AHP와 SIT의 결합 비교

(표 8)은 EFAM/Fuzzy AHP에서 얻어진 개념의 최종 중요도와 다양한 SIT들과 결합하여 개념들에 대한 분류값들의 순위를 감성기준을 고려하여 비교하여 나타내었다.

감성기준 고려에 따른 분류값1은 EFAM/Fuzzy AHP와 변환에 의한 유사도로 계산하였고 분류값2는 EFAM/Fuzzy AHP와 min 연산자를 사용하였다. 분류값3은 EFAM/Fuzzy AHP와 디스크 계수에 의하여 구하였고, 분류값4는 EFAM/Fuzzy AHP와 코사인 유사도를 사용하였다. 분류값5는 EFAM/Fuzzy AHP와 확장 자카드 계수 유사도를 사용하여 계산하였다.

EFAM/Fuzzy AHP에서 분류값1, 분류값3, 분

류값4, 분류값5는 감성기준을 고려했을 경우와 고려하지 않았을 경우 동일한 결과를 얻었지만 분류값2는 ‘정전’과 ‘네트워크 부하’의 순위가 다름을 알 수 있었다. ‘해커’의 경우에 분류값1 순위는 10위로 EFAM/Fuzzy AHP에서는 6위를 기록하여 순위의 차가 4로 가장 많은 순위차가 발생하였다. 다음으로 순위의 차가 2가 나는 ‘프로그램 버그’와 ‘네트워크 부하’가 있고 그 다음은 순위의 차가 1씩 발생한 개념 네 개가 확인되었다.

개념의 중요도를 순위별로 나타낸 결과 다른 방법들은 모두 감성기준 고려와 상관없이 동일한 결과를 얻었지만 분류값2만 감성기준을 고려 여부에 따라 ‘정전’과 ‘네트워크 부하’의 순위가 변하는 것을 관찰할 수 있었다.

(표 9) 문제 영역에서 문서추출 비교

결합한 분류 방법	추출 문서수	비율
EFAM + 변환에 의한 유사도	27,300	0.2929%
EFAM + min 연산자(감성기준 고려안함)	2,900	0.0311%
EFAM + min 연산자(감성기준 고려)	2,880	0.0309%
EFAM + 다이스 계수	2,970	0.0319%
EFAM + 코사인 유사도	2,960	0.0318%
EFAM + 확장 자카드 계수	2,980	0.0320%

(표 9)는 ‘웹서버에서 발생할 수 있는 고장 원인’의 특정 문제 영역에서 EFAM과 각 SIT들을 결합한 방법들을 문서추출에 적용해 본 결과를 나타내었다.

(표 9)에서 나타내었듯이 ‘웹서버에서 발생할 수 있는 고장’ 문제 영역에 대해 Google에서 총 9,320,000개의 문서들 중에서 EFAM과 각 SIT 방법들을 결합한 개념 분류값의 순위에 따라 검색한 결과 감성기준을 고려한 EFAM과 min 연산자의 EFASIT 모델 결합이 비교적 효율적으로 문서 검색을 하였다는 것을 알 수 있었다.

4.5 퍼지지식 통합과 규칙 생성

‘웹 서버에서 발생할 수 있는 고장’에 사용할 일곱 가지의 가능한 의사결정을 추출한다. 즉 ‘DB 서버 파괴’, ‘DB 응답 느림’, ‘서버 파괴’, ‘서버 응답 느림’, ‘네트워크 고장’, ‘사용자는 웹페이지 접근 불가’, ‘애플리케이션 이상’ 등이다.

다수 전문가 e_1, e_2, e_3 로부터 도출된 개념과 객체 그리드에서 사용되는 확신도는 0에서 1사이의 값을 사용하고 퍼지 값은 7점 척도로 -3에서부터 3까지의 정수로 나열된다. 이것은 언어형태와 퍼지값의 적절한 조합으로 나타낸다. 퍼지 값 -3은 ‘매우 낮음’, ‘매우 짧음’, 또는 ‘매우 느림’, -2는 ‘낮음’, ‘짧음’, 또는 ‘느림’, -1은 ‘다소 낮음’, ‘다소 짧음’ 또는 ‘다소 느림’, 0은 ‘평균’, ‘중간’, 또는 보통 1은 ‘다소 높음’, ‘다소 김’, 또는 ‘다소 빠름’ 2는 ‘높음’, ‘김’ 또는 ‘빠름’, 3은 ‘매우 높음’, ‘매우 김’ 또는 ‘매우 빠름’을 의미한다.

개념과 객체 그리드를 가지고 통합된 cf와 fv 값을 기반으로 퍼지 규칙1에서 퍼지 규칙 7까지의 의사결정을 생성하였다. 생성한 7개의 규칙 중 예로 퍼지 규칙1을 표시하면 다음과 같다.

■ 규칙 1 - DB 서버 파괴

$$\begin{aligned}
 & [dcf_1/dfv_1, dcf_2/dfv_2, \dots, dcf_k/dfv_k, \dots, dcf_m/dfv_m] \\
 & = [0.32, 0.33, 0.35] \\
 & \cdot \begin{bmatrix} 0.9/2, 0.8/2, 0.7/3, & 0.7/1, & 0.9/3, & 0.5/0, & 0.8/3, 0.6/-2, 0.5/-3, 0.6/1 \\ 0.9/2, 0.8/2, 0.5/3, & 0.8/2, & 0.9/3, & 0.6/2, & 0.8/3, 0.7/-1, 0.5/-3, 0.8/1 \\ 1/3, & 0.9/3, & 1/3, & 0.7/-1, & 0.9/3, 0.5/-2, 0.9/3, 0.6/-2, 0.5/-3, 0.8/1 \end{bmatrix} \\
 & = [0.935/2.35, 0.835/2.35, 0.739/3, 0.733/0.63, 0.9/3, \\
 & 0.533/-0.04, 0.835/3, 0.633/-1.67, 0.5/-3, 0.74/1] \\
 & cf = [0.935, 0.835, 0.739, 0.733, 0.9, 0.533, 0.835, 0.633, 0.5, 0.74] \\
 & \cdot [0.127, 0.141, 0.222, 0.080, 0.061, 0.114, 0.061, 0.063, 0.065, 0.068] \\
 & = 0.747 \\
 & fv = Round[2.35, 2.35, 3, 0.63, 3, -0.04, 3, -1.67, -3, 1] \\
 & = [2, 2, 3, 1, 3, 0, 3, -2, -3, 1]
 \end{aligned}$$

다음은 통합된 cf와 fv 값을 기반으로 퍼지 규칙을 생성하였다.

규칙 1 : IF 하드디스크 공간이 충분하고
 메모리 공간이 충분하고
 컴퓨터바이러스의 위험이 매우 높고
 하드웨어 손상은 다소 심각하고
 정전은 매우 심각하고
 프로그램 버그가 존재하고
 운영체제 고장은 매우 심각하고
 CPU 부하는 낮고
 네트워크 부하도 매우 낮으며
 해커 침입 경도가 다소 높으면

THEN 시스템 상태는 $cf = 0.747$ 로 ‘DB 서버 파괴’이다.

5. 결 론

본 논문에서는 의사결정에 관한 요소를 계층화 구조로 나타내는 AHP나 Fuzzy AHP 방법들을 데이터의 관점에서 대안, 평가 기준, 주관적 속성 가중치, 개념과 객체 사이에 퍼지 관계를 기반으로 더 효과적으로 개선하여 웹 자원을 관리하고 의사결정을 할 수 있는 EFASIT 모델을 제안하였다.

EFAM은 다수 전문가로부터 개념을 획득하여 공통 개념 그리드를 구성하고 전문가별 과거 경험, 배경지식, 관련평판 및 감성에 대한 기준을 가지고 자기-특점을 구하여 전문가의 가중치를 통합하여 EFAM, Fuzzy AHP 방법을 사용하여 개념 중요도를 비교하였고 개념들의 최종 개념가중치를 계산하였다. SIT는 개념들 간의 근접관계를 고려하기 위하여 다수 전문가들에 의해 통합된 개념 유사도 와 소속도로 구분하여 변환에 의한 유사도, 퍼지 집합의 min 연산자, 다이스 계수, 코사인 유사도, 확장 자카드 계수를 사용하여 계산하고 분류하였다.

EFAM과 SIT를 결합한 EFASIT 모델은 개념들에 대한 순위 비교를 분석하여 문서를 검색한 결과 감성 기준을 고려한 EFAM과 min 연산자의 EFASIT 모델 방법이 다른 EFASIT 모델 방법들보다 비교적 우수함을 알 수 있었다. 그리고 개념과 객체간의 의사결정을 하기 위해 퍼지 지식을 통합하고 새로운 규칙들을 생성하였다.

따라서 EFASIT 모델은 웹상의 효율적인 문서 검색과 특정 영역의 문제를 진단하기 위하여 영역 코퍼스로부터 추출된 개념들이 가지는 관계의 계층화와 의미론적 내용을 바탕으로 EFAM과 각 SIT를 결합하여 문서를 추출하는 방법을 제안하고 퍼지지식의 통합으로 도출된 규칙들은 문제에 대해 더 명확한 의사결정을 할 수 있음을 알 수 있었다.

제안한 모델은 온톨로지 학습 애플리케이션, 의사결정지원 시스템, 자연어 처리, 텍스트 마이닝, 시스템의 고장 진단 및 실용적인 지식 모니터링 시스템 툴 개발 등에 광범위하게 적용될 수 있을 것이다.

향후 연구과제로는 특정 분야에 대한 온톨로지를 여러 분야의 전문가들이 공동 작업을 통해서 설계하여 기존의 통계 및 규칙기반의 텍스트 마이닝 기술에 온톨로지 기반 지식 검색 기능을 부여하여 한 단계 발전된 정보 추론을 수행하는 연구가 요구된다.

참 고 문 헌

- [1] Naoya Kotami and Yukio Kodono, "A Study on AHP using Fuzzy Theory," 20th Fuzzy System Symposium, 2004.
- [2] Tien-Chin Wang and Yueh-Hsiang Chen, "Applying Fuzzy Linguistic Preference Relations to the Improvement of Consistency of Fuzzy AHP", Information System, Vol. 178, pp. 3755-3765, 2008.
- [3] 정환목, "다치논리함수를 이용한 감성처리 모델," 한국지능시스템학회 논문지, Vol. 19, No. 1, pp. 13-18, 2009.
- [4] C. S. Son and H. M. Chung, "An Emotion Classification Based on Fuzzy Inference and Color Psychology," *International Journal of Fuzzy Logic and Intelligent Systems*, Vol. 4, No. 1, pp. 18-22, 2004.
- [5] 류경현, 정환목, "MFAC를 사용한 근접관계의 분류," 한국 퍼지 및 지능시스템학회 논문지, Vol. 18, No. 1, pp. 139-144, 2008.
- [6] Supriya Kumar De, Ranjit Biswas, and Akhil Ranjan Roy, "On Extended Fuzzy Retional Database Model with Proximity Relations," *Fuzzy Sets and Systems*, Vol. 117, pp. 195-201, 2001.

- [7] J. Mishra and S. D. Ghosh, "A Study of Fuzzy Relational Database," *International Journal of Computational Cognition*, Vol. 6, No. 4, pp. 45-50, 2008.
- [8] Hsuan-Shih Lee, "On fuzzy preference relation in group decision making", *International Journal of Computer Mathematics*, Vol. 82, Issue 2, pp. 133-140, 2005.
- [9] Hai Dong, Farookh Khadeer Hussain and Elizabeth Chang, "A Hybrid Concept Similarity Measure Model for Ontology Environment", in Robert Meersman, Pilar Herrero and Tharam Dillon (ed), *On the move to meaningful internet systems: OTM 2009 workshops*, pp. 848-857, Springer-Verlag, 2009.
- [10] 노상규, 박진수 공저, "온톨로지 웹2.0에서 3.0으로", 가즈토이, 2007.
- [11] Rachel Conrad, California power outages suspended-for now, CNET news.com, <http://news.cnet.com/2100-1017-251167.html>, 2001.
- [12] Craig Newell, Network Card Theft Causes Internet Outage, <http://www.eweek.com/c/a/Security/Network-Card-Theft-Causes-Internet-Outage>, 2004.

◎ 저 자 소 개 ◎

류 경 현



1990년 대구가톨릭대학교 전산통계학과 졸업(학사)
1992년 대구가톨릭대학교 대학원 전산통계학과 졸업(석사)
2009년 대구가톨릭대학교 대학원 컴퓨터정보통신공학과 졸업(박사)
2009년~현재 대구대학교 시간강사
관심분야 : 소프트웨어공학, 지능형시스템, 의사결정, 온톨로지 etc.
E-mail : ryu0919@naver.com

피 수 영



1987년 대구가톨릭대학교 통계학과 졸업(학사)
1989년 대구가톨릭대학교 대학원 전산통계학과 졸업(석사)
2000년 대구가톨릭대학교 대학원 컴퓨터정보통신공학부 졸업(박사)
2000년~현재 대구가톨릭대학교 실용컴퓨터 책임교수
관심분야 : 데이터마이닝, 지능형시스템, 의사결정, 확률추론 etc.
E-mail : sypi@cu.ac.kr