

온라인 뉴스 가치 평가 및 개인화 기법

최광선¹, 김수동^{2*}

¹솔트룩스 CDS사업본부, ²송실대학교 컴퓨터학부

A Method for Evaluating Online News Value and Personalization

Kwang Sun Choi¹, Soo Dong Kim^{2*}

¹CDS Business Center, Saltlux Inc.

²Department of Computer Science, Soongsil University

요약 본 논문의 연구 목표는 뉴스 가치 평가에 근거한 중요 뉴스 자동 추천 및 개인화 방안을 제시하는 데에 있다. 뉴스 가치 평가는 전통적인 오프라인 신문에서 편집자들이 1면 뉴스를 선정할 때 사용하는 접근법으로 본 논문에서는 이를 시스템적으로 구현하는 방안을 제시한다. 이렇게 함으로써 콘텐츠 주제에 대한 전통적인 개인 선호 성향과는 다르게 뉴스의 사회적 가치에 대한 관심 성향을 기준으로 중요 뉴스를 선별할 수가 있다. 뉴스의 사회적 가치는 지면 신문의 기존 연구에서 제시한 사회적 중요도, 새로운 볼거리, 수용자 관련성, 인간적 흥미 4가지 기준을 준용하였고, 본 연구에서는 이를 시스템적으로 적용하기 위한 절차적, 구조적 방안을 도출하였다. 중요 뉴스의 선별 과정은 뉴스의 가치 평가를 위한 과정과 평가된 결과를 개인화하는 과정으로 구성된다. 실험을 통해 특정 시점에서의 각 온라인 뉴스 서비스들의 중요 뉴스들과 본 논문에서 제시한 기법을 통해 선별된 중요 뉴스들에 대한 사용자 만족도를 비교 평가하여 본 연구에서 제안하는 방법이 더 효과적임을 확인하였다.

Abstract The purpose of this paper is to propose a method for recommendation and personalization of important news articles based on evaluating news value. Evaluation of news is the approach by which editors select news articles for cover-story in traditional offline news papers area. For this, my study proposes a suite of methods to select and personalize a set of news based on evaluating news articles, not just on the personal preference for them. The aforementioned the value of news articles including social impact, novelty, relevance to each audience, and human interest, all of which have been factorized in many previous studies, is a main concept for a procedural and structural application methodology deduced in this study. After a comparative case study with other online news services, it was shown that my research provides more effective way to select important news articles in terms of user satisfaction than others.

Keywords : Content Recommendation, Online News Valuation, News Article Recommendation, Consumer Preference, Personalizaion

1. 서론

뉴스 서비스는 오랜 기간 동안 지면 신문과 TV나 라

디오와 같은 전통적인 매체들을 통해 제공되어온 지식 서비스이다. 최근 들어, 지식 소비환경이 오프라인 중심에서 온라인으로 중심으로 진화함에 따라 뉴스 서비스의

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임.

(No. R0101-15-0054, WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발)

*Corresponding Author : Soo Dong Kim(Soongsil University)

Tel: +82-2-820-0909 email: sdkim777@gmail.com

Received July 8, 2015

Revised (1st September 1, 2015, 2nd September 18, 2015)

Accepted December 4, 2015

Published December 31, 2015

소비환경도 자연스럽게 온라인을 중심으로 진화하였다. 그리고 이러한 과정에서 네이버, 다음, 구글, 줌닷컴과 같은 검색 포털 서비스들은 각 언론사의 뉴스들을 모아 서 통합적으로 제공하는 서비스를 제공하게 되었다. 또한 뉴스는 멈추지 않고 산되고 소비되는 뉴스의 양 역시 지속적으로 증가하고 있으며, 이러한 뉴스에 대한 중요성도 지속적으로 높아지고 있다.

반면에 증가되는 뉴스의 양은 뉴스 소비자들에게 원하는 뉴스를 찾기 위해 더 많은 노력을 들일 것을 요구하고 있다. 게다가 뉴스 소비 환경의 중심이 데스크톱에서 모바일로 이동하고 있어서 뉴스 소비자들이 뉴스를 발견하여 소비하는 데 사용되는 시간은 점점 더 짧아지고 있다. 특히 다른 콘텐츠들에 비해 상대적으로 양이 적은 뉴스의 특성은 이러한 소비 패턴의 변화를 더욱 가속화 시키고 있다.

콘텐츠 소비자로 하여금 보다 짧은 시간에 콘텐츠를 발견할 수 있도록 돕는 기술적 해결책은 추천 방법을 이용하는 것이다. 추천 방법은 뉴스를 포함하여 모든 온라인 지식 및 콘텐츠 서비스 분야에서 광범위하게 연구되고 적용되고 있는 전통적인 주제이다.

대표적인 추천 방법으로는 콘텐츠의 인기도를 측정해서 해당 인기도를 기반으로 불특정 다수에게 콘텐츠를 추천하는 인기 기반 추천 방법이 있다. 한편 특정 개인에게 맞춤형으로 추천하는 방법으로는 추천 대상자가 소비한 콘텐츠들의 특성과 유사한 콘텐츠를 추천하는 콘텐츠 기반 필터링(content based filtering)과 추천 대상자와 유사한 콘텐츠 소비 성향을 가지고 있는 사람들이 소비한 다른 콘텐츠를 추천하는 협업 필터링(collaborative filtering), 그리고 이러한 방법들을 혼합한 혼합 필터링(hybrid filtering) 등의 방법이 있다.

온라인 뉴스 서비스에서도 이러한 추천 방법들을 이용하여 서비스를 제공하고 있다. 인기도 기반 추천 방법의 경우, 인기도를 계산하기 위해서 주로 ‘뉴스를 읽은 수’라는 특성(feature)을 이용하는 데, 몇 명의 사람들이 해당 뉴스를 읽었는지를 기준으로 인기도를 측정한다. 콘텐츠 기반 필터링의 경우, 추천 대상자가 주로 읽은 뉴스의 카테고리 또는 해당 뉴스들이 가지고 있는 키워드들을 특성으로 추출하고 해당 특성을 이용해서 추천 대상자가 주로 읽는 뉴스와 유사한 뉴스나 같은 분야의 뉴스를 추천할 뉴스로 선별한다. 협업 필터링의 경우, 추천 대상자와 읽은 뉴스를 특성으로 간주하고 추천 대상자와

유사한 성향의 소비자를 선별한 후에 유사한 성향의 소비자들끼리 함께 읽은 뉴스들을 추천할 뉴스로 선별한다.

하지만 위에서 설명한 뉴스 추천 기법들은 여러 가지 한계점들을 가지고 있다. 예를 들어, 사람들이 뉴스를 읽은 수를 이용한 인기도 기반 추천 기법의 경우 새로운 뉴스가 추천될 가능성은 상대적으로 적다. 콘텐츠 기반 필터링의 경우, 추천 대상자가 주로 읽는 뉴스의 카테고리 또는 키워드를 기반으로 추천하기 때문에 주제가 유사한 뉴스만 추천되는 문제가 발생한다(새로운 소식 중에서 중요한 소식들을 전달해야 하는 뉴스 속성에 적합하지 않다). 협업 필터링의 경우도 추천 대상자와 유사한 성향의 사람들이 읽은 뉴스를 추천하기 때문에 해당 뉴스 서비스를 이용한지 오래되지 않은 사용자의 경우 뉴스를 추천되지 못하는 문제가 발생한다. 이러한 문제를 콜드 스타트(cold start)라고 한다. 또한, 협업 필터링의 경우, 추천 대상자와 유사한 사용자를 찾기 위한 계산 비용이 크기 때문에 빠르게 소비되는 뉴스 추천에서 이용하기 어렵다. 이러한 문제점은 추천 방법 자체가 가지고 있는 문제점이기도 하지만 해당 추천 방법에서 사용하고 있는 특성들이 가지는 한계이기도 하다. 즉, 현재의 자동화된 추천에서 사용하는 특성들과 현재의 적용 방법으로는 새로운 뉴스를 추천하거나, 다양한 주제의 중요한 뉴스를 기 어렵다가나 다양한 주제의 뉴스를 추천하는 데에 한계가 있는 것이다.

반면에 전통적인 지면 신문에 있어서 1면 기사는 신문의 편집자들이 신문 구독자 즉, 뉴스의 소비자들로 하여금 중요한 기사들을 빠르게 발견하고 소비할 수 있도록 도와주는 핵심적인 추천 서비스이다. 특히 1면에 게재할 뉴스를 선정하는 것은 각 신문사들의 특징을 결정짓는 핵심 활동으로 각 신문사의 편집자들이 갖는 고유한 권한이자 핵심 역량이었다. 그래서 언론계에서는 오랜 기간 동안 이를 체계화하고 객관적인 선별 기준을 만들기 위해 다양한 노력을 해왔다.

그런데 이러한 전통적인 지면 신문에서의 연구 결과들에서 중요한 뉴스를 선별하기 위해 이용하는 특성들을 살펴보면 단지 뉴스의 주제어나 주제 분류 혹은 인기도 등의 단순한 특성들만이 아닌, 소비자들이 뉴스를 어떻게 소비하는가와 관련한 활용 가치 관점의 특성을 이용한다는 특색이 있다. 뉴스가 소비자에게 제공할 수 있는 활용적 가치들은 사회적인 중요도(social significance), 새로운 볼거리(visual novelty), 사용자 관련성(audience

relevance), 인간적 흥미(Human Interest) 등으로 분류될 수 있는데, 이러한 접근법은 온라인 뉴스 서비스의 자동화된 추천에서도 유용할 것으로 판단된다.

본 논문에서는 위에서 설명한 뉴스의 활용적 가치에 해당되는 다양한 특성들을 이용하여 온라인 뉴스의 자동화된 추천에 적용하는 방법과 시스템을 제안한다. 또한, 이러한 요소들을 이용해서 불특정 다수의 뉴스 소비자들을 대상으로 뉴스를 추천하는 방법을 제안하며, 이를 개인화하기 위해 콘텐츠 기반 필터링에 이러한 특성들을 적용하는 방법을 제안하려고 한다.

제안되는 추천 방안에서는 뉴스의 카테고리 또는 주제어와 같이 단순한 뉴스의 내용적인 특성들만을 활용하는 한계를 벗어나기 위해 추천 대상이 되는 뉴스들로부터 의미적 정보 요소들을 추출하고, 의미적 정보 요소들에 근거하여 뉴스의 활용적 가치를 측정하며, 이를 기준으로 중요 뉴스를 선별한다. 또한, 이러한 과정에서 해당 뉴스의 생산자와 소비자가 속해있는 다양한 사회 구성원들과 이들 간의 관계와 피드백을 활용하고자 한다. 궁극적으로는 이러한 접근을 통해 새로 생성된 뉴스의 대해 객관적 근거를 가지고 중요한 뉴스에 선정될 수 있는 기회를 보장하거나, 개인화된 추천의 경우에도 소비자의 과거 이력에 편향되지 않는 다양한 추천이 가능하도록 하여 추천 결과에 대한 소비자의 만족도를 향상시킬 수 있다.

본 연구에서 제안하는 방안의 효과를 검증하기 위해서는 실제 모바일에서 서비스되고 있는 뉴스들을 대상으로 제안 방안을 적용하여 추천한 뉴스와 현재 서비스되고 있는 각 포털 서비스에서 추천하고 있는 뉴스를 특정 사람들에게 제공하여 만족도를 확인한다. 또한, 제안한 방안에서 이용하는 특성들이 개인화에도 적용 가능한지 확인하기 위해서 불특정 다수를 대상으로 하는 제안 방안과 개인화 방법을 적용한 제안 방안의 추천 뉴스에 대한 만족도를 비교한다. 실험결과, 제안하는 방안이 각 포털 서비스에서 제공하는 뉴스 추천보다 사용자 만족도가 높았으며 개인화 추천 기법을 적용했을 때도 본 논문에서 제안한 요소들이 의미 있다는 것을 확인했다.

본 논문의 구성은 다음과 같다. 먼저, 2장 관련 연구에서는 전통적인 온라인 콘텐츠 추천 방법들과 지면 신문에서의 1면기사 선정과 관련한 연구들의 특징을 고찰하고, 3장에서는 관련 연구에서 고찰된 오프라인 뉴스를 위한 가치 평가 모델을 온라인 뉴스 서비스에 적용하기

위한 절차와 기법을 제시하며, 4장에서는 뉴스 가치 평가 모델을 개인화하기 위한 절차와 기법을 제안한다. 마지막으로 5장에서는 실험적 환경 구성을 통해 뉴스 가치 평가 모델 기반의 추천 방안을 검증하고, 6장에서 결론을 맺고자 한다.

2. 관련연구

온라인 뉴스 서비스에서 뉴스의 추천은 매우 전통적인 주제이며 다양한 연구가 진행되었다. 이러한 연구들에 있어서 전통적으로는 크게 뉴스의 내용을 중심으로 추천을 하는 콘텐츠 기반 필터링(Content based Filtering) 방식과 소비자 관심을 중심으로 추천을 하는 협업 필터링(Collaborative Filtering) 방식 그리고 이 두 가지가 혼합된 방식(Hybrid Filtering) 등이 연구되었다.

내용 중심의 추천 방식은 뉴스의 내용을 구성하는 단어, 주제어, 혹은 주제 분류 등을 중심으로 내용적 특성을 추출하고, 이들 특성들 간의 유사성을 이용하여 내용적으로 유사한 뉴스를 선별하여 추천하는 방식이다. 이러한 연구들에서는 단어들 간의 의미적 네트워크를 기반으로 뉴스들을 군집하는 방법[3], 주제어 간의 확률적 관계모델인 주제어 모델을 이용하여 군집하는 방법[4], 새롭게 게재되는 최신 뉴스들의 스트림을 시간적 제약 안에서 분석하여 추천하는 방법[5] 등이 제시되었다.

소비자 관심 중심의 추천 방식은 소비자들의 뉴스 소비 행위 이력과 행위의 대상이 되는 뉴스들을 중심으로 소비자들의 관심에 대한 프로파일을 구성하는 내용적 특성 추출하고 해당 특성에 적합한 뉴스를 추천하는 방식이다. 이때 특성은 역시 단어나 주제어 혹은 주제 분류 등을 중심으로 구성된다. 이러한 연구들에서는 토픽모델을 중심으로 소비자 프로파일의 특성을 추출하여 이용하는 방법[4], 웹 서버상의 뉴스 소비 이력을 분석해서 소비자의 프로파일을 추출하여 이용하는 방법[6], SNS인 트위터에서의 소비자의 행위 이력을 이용하여 소비자 프로파일을 추출하여 이용하는 방법[7], 스마트폰에서의 뉴스 소비자 이력 분석을 통해 소비자 관심을 파악하여 이용하는 방법[8] 등이 제시되었다.

그런데 이러한 전통적인 방식들은 공통적으로 뉴스의 제목과 내용으로부터 추출된 단어나 주제어, 주제 분류 등 언어 형태적인 요소들을 기준으로 소비자의 선호 특

성을 학습한다. 그러므로 이러한 방식은 기사 내용에 대한 소비자의 선호가 언어 형태적으로 식별되었을 때 매우 효과적인 방식이다. 하지만 소비자가 특정 기사를 읽기 위해 선택하는 이유를 기사 내용에 대한 선호만이 아니고 단정하는 것은 불완전하다[9].

한편 언론학 분야에서는 뉴스 소비자들의 뉴스 소비 성향과 관련하여 뉴스 가치를 기준으로 뉴스 소비자에게 뉴스를 추천하는 기준에 대한 연구를 진행하여 왔다. 특히 이종혁 등의 연구는 해당 주제의 많은 연구들을 비교 분석하여 하나의 통합된 기준을 제시하고 있다. 여기서 뉴스 가치란 '어떤 사건이 보도될 만한 가치를 가지는 가라는 질문에 대한 개념적 응답으로서, 해당 연구에서는 많은 언론 전문가들의 의견을 수렴해서 객관화시켜 집대성하였다. 이러한 연구들은 뉴스 소비자들이 뉴스를 선택하는 기준이 특정 뉴스 주제나 분야를 좋아하는 것과는 별개의 다른 관점들의 가치들이 있다는 것을 제시하고 있다. 이러한 방식은 온라인 뉴스 서비스의 1면 기사를 선정하여 추천하는 과정에서 매우 유용해 보인다. 하지만 해당 연구는 개념적 가치 기준을 제시할 뿐 이러한 가치 기준을 이용한 자동화된 뉴스 가치 기반의 추천 방법을 제시하고 있지는 못하다[2].

그러므로 온라인 뉴스 서비스의 1면기사 추천의 품질을 향상시키기 위해서는 전통적인 뉴스 추천 방법을 활용하는 뉴스 기사의 내용에 대한 선호 특성만을 이용한 추천뿐만 아니라, 뉴스의 가치를 기반으로 한 다양한 관점에 대한 선호 특성을 이용하는 것이 중요하다[10].

3. 온라인 뉴스 가치 평가 절차 및 기법

지면 신문에서는 발간하는 신문의 판매를 확대하기 위해 좋은 기사를 작성하는 것 외에도 신문의 1면기사(커버스토리, Cover Story)로 노출시킬만한 가치 있고 중요한 기사를 선별하고, 중요한 기사들이 더욱 눈에 잘 띄도록 배치하는 것이 매우 중요하다. 통상 이러한 작업은 편집자(데스크)를 중심으로 이루어지며 중요한 기사를 선별하기 위한 편집자의 역량은 신문 판매에 절대적인 영향을 끼치게 된다.

그러나 지면 신문의 뉴스 기사에 대한 가치 평가는 매우 감각적이고 주관적인 의사결정의 과정으로 다분히 편집자의 기준에 크게 좌우될 수밖에 없다. 이러한 특징으

로 인해 1면기사의 기사 구성의 만족도는 항상 일정할 수가 없다. 이러한 이유로 언론 분야에서는 오랫동안 기사의 가치 결정을 위한 객관적인 기준을 정립하기 위한 연구를 해왔다. 특히 이종혁 등의 연구에서는 기존 다양한 관련 문헌들로부터 56개의 뉴스 가치 개념들을 수집해 정리하였고, 이들 간의 관계를 분석해 4개의 1차 요인(최상위 가치요인), 10개의 2차 요인(차상위 가치요인), 30개의 하위 측정항목들로 구성된 '뉴스 가치 구조 모델'을 제시하였다. 또한, 제시된 모델에서 최상의 요인으로서의 사회적 중요도(social significant), 새로운 볼거리(visual novelty), 수용자 관련성(audience relevance), 인간적 흥미(human interest) 등을 제시하고 있으며, 각각의 차상위 가치요인들을 기준으로 가중치를 설정하여 뉴스 기사에 대한 가치 평가의 틀을 제시하였다. 아래 Table 1은 위 연구에서 제시한 '뉴스 가치 구조 모델'이다 [2].

Table 1. News Value Structure Model

Top Level Factor	Second Level Factor	Second Level Factor Weight	Low Level Factor
Social Significance	Impact	.72	Opinion Leader Impact, Issue Relevance, Article Relevance
	Prominence	.56	Well Known People Influence, Political Importance
	Depth	.85	Depth of Event, Subsequent Event Induction, Interaction Induction, Participation Induction
	Conflict	.62	Conflict Structure, Conflict Intensity, Conflict Valance
Visual Novelty	Novelty	.51	Statistical Deviance, Novelty, Unexpected News
	Action	.83	People Activity, Visual Content, Visual Effect
Audience Relevance	Usefulness	.99	Facts, Statistics, Aggregates, Publicity
	Proximity	.57	Physical Proximity, Psychological Proximity, Cultural Proximity
Human Interest	Entertainment	.89	Entertainingness, Humor
	Story	.50	Story Centric, People Centric, Drama

그러나 구성된 뉴스 가치 평가를 위한 기준들을 보면, 전통적인 콘텐츠 추천 시스템에서 이용하는 콘텐츠 주제 분류에 대한 콘텐츠 소비자의 선호 특성을 이용하는 것과는 다소 다른 가치 평가 요소들을 발견할 수 있다.

우선 최상위 가치요인으로 사회적 중요도(Social Significance), 새로운 볼거리(Visual Novelty), 사용자 관련성(Audience Relevance), 인간적 흥미(Human Interest) 등 4가지 요인을 살펴보면 뉴스 기사의 주제나 내용에 대한 단순한 선호가 아닌, 뉴스 기사가 뉴스 소비자에게 제공하는 의미적 가치를 기준으로 함을 알 수 있다.

본 연구에서는 온라인 뉴스 서비스의 중요 기사 선별을 자동화하기 위해 Table 1.에 제시된 '뉴스 가치 구조 모델'에 따라 뉴스 기사의 가치를 평가하여 뉴스 기사의 중요도인 순위점수(Ranking-Score)형태로 측정한다.

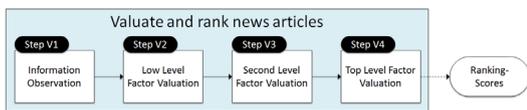


Fig. 1. Valuation Process for News Articles

자동화된 뉴스 가치 평가의 절차는 Fig. 1 과 같이 Step V1. 정보 관찰(Information Observation), Step V2. 하위 가치요인 평가(Low Level Factor Valuation), Step V3. 차상위 가치요인 평가(Second Level Factor Valuation), Step V4. 최상위 가치요인 평가 등으로 구성된 총 4단계의 절차로 진행된다.

Step V1. 정보 관찰 단계에서는 자동화된 정보 수집 기술을 이용하여 관찰 가능한 정보들을 획득하고 분석에 이용할 수 있는 정형화된 형태로 변환하여 적재한다.

Step V2. 하위 가치요인 평가 단계에서는 각의 뉴스 기사를 하위 가치요인에 해당하는 특성들을 식별하고, 식별된 특성들의 빈도 등을 이용해서 하위 가치요인에 대한 평가점수(LLF Value)를 계산한다.

Step V3. 차상위 가치요인 평가 단계에서는 하위 가치요인 평가 결과와 각각의 하위 가치요인들에 대한 하위 가치요인 가중치 모델을 적용하여 차상위 가치요인에 대한 평가점수(SLF Value)를 계산한다.

Step V4. 최상위 가치요인 평가 단계에서는 뉴스 기사에 대한 통합적인 가치를 평가하는 최종 단계로서, 차상위 가치요인에 대한 평가점수에 차상위 가치요인 가중치 모델을 적용하여 최상위 가치요인에 대한 평가점수

(TLF Value)를 계산한다.

3.1 Step V1: 정보 관찰

정보 관찰 단계에서는 자동화된 정보 수집 기술을 이용하여 관찰 가능한 정보들을 획득하고 분석에 이용할 수 있는 정형화되어 분석 가능한 형태로 변환하여 적재한다. 관찰 가능한 정보는 이용 가능한 기술을 통해 자동으로 정보 접근과 수집이 가능한 정보들을 말하며, 관찰 가능 정보의 종류가 다양해질수록 평가할 수 있는 하위 가치요인의 수도 다양해질 수 있다.

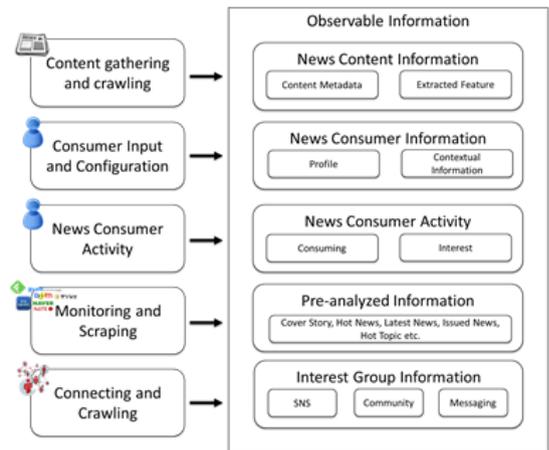


Fig. 2. Observable Information for Evaluating Online News Value

Fig. 2 과 같이 관찰 가능한 정보(Observable Information)는 뉴스 콘텐츠 정보(News Content Information), 뉴스 소비자 정보(News Consumer Information), 뉴스 소비자 행위(News Consumer Activity), 사전 분석 정보(Pre-analyzed Information), 관심 그룹 정보(Interest Group Information) 등 5가지의 정보가 있다.

1) 뉴스 콘텐츠 정보(News Content Information)

뉴스의 메타데이터로부터 관찰되는 정보들(Content Metadata : 제목, 게시일시, 기자 혹은 저자, 언론사 혹은 출처, 주제 분류 등)과 뉴스의 내용을 분석하거나 뉴스의 내용으로부터 자동 혹은 수동으로 추출할 수 있는 정보들(Extracted Feature : 관련인물, 관련사건, 관련 장소, 관련기

관, 관련시점 등)

- 2) 뉴스 소비자 정보(News Consumer Information)
온라인 뉴스 서비스 제공자의 개인정보 수집 수준에 따라 매우 다양하게 구성될 수 있으며, 소비자 개인이 제공하는 개인정보(Profile : 나이, 성별, 직업, 관심지역, 관심주제어), 서비스 이용과정에서 관찰되어 제공될 수 있는 상황정보(Contextual Information : 현재시간, 현재위치)등에 해당됨
- 3) 뉴스 소비자 행위(News Consumer Activity)
뉴스 소비자 개개인이 뉴스를 소비하는 과정에서 관찰되는 각각의 행위들로 구성되며, 크게 단순히 뉴스를 발견하고 소비하는 행위들(Consuming : 탐색, 검색, 목록조회, 내용읽기 등) 특정 뉴스에 대한 개인의 관심을 표현(Interest : 공유, 댓글, 평가, 스크랩 등)하는 행위로 구분됨. 해당 기사의 사회적 관심도를 측정할 때 활용됨
- 4) 사전 분석 정보(Pre-analyzed Information)
온라인 뉴스 서비스를 제공하는 과정에서 이미 분석되어 제공된 순위 정보들(Cover Story, Latest News, Hot News, Issued News, Hot Topic 등). 해당 기사의 사회적 중요도를 측정할 때 활용됨
- 5) 관심 그룹 정보(Interest Group Information)는 SNS, 커뮤니티(Community), 메시징(Messaging) 서비스처럼, 뉴스 소비자들 상호간의 교류 속에서 행해지는 뉴스 소비 행위의 관찰을 통해 획득됨. 특히 이렇게 연결된 관계 그룹들이 뉴스 소비자 개개인에게 큰 영향을 미치는 상호작용하는 그룹이라는 점에서 뉴스 가치 측정을 위한 중요한 정보로 활용됨

관찰 가능한 정보들에 대한 접근 방식과 수집 방식은 각각의 정보 출처에 따라 다르다. 본 연구에서는 크게 다음과 같은 3가지 법을 이용하여 수집하였다.

- 1) 웹 크롤링 및 정보 추출
표준 HTTP 프로토콜을 이용하여 웹상에서 수집함. 또한 수집된 웹 콘텐츠로부터 제목, 게재일자, 저자 등 필요한 속성값을 추출함. 웹상의 노출되는 모든 정보들을 대상으로 수집할 수 있음.
- 2) RSS 크롤링
RSS형태로 제공되는 뉴스 기사들의 경우에는

RSS를 별도로 크롤링해서 뉴스 기사 목록을 확보한 뒤, 해당 기사를 웹 크롤링을 통해 수집함

- 3) 정보 제공 API
페이스북, 트위터 등과 같이 정보 제공을 위한 API(보통은 Open API 제공)를 이용하여 정보를 수집함. 단순한 콘텐츠 외에 관련한 메타데이터를 함께 확보할 수 있음

3.2 Step V2: 하위 가치요인(LLF) 평가

하위 가치요인(Low Level Factor, LLF)은 관찰 가능한 정보들을 이용하여 직접적으로 측정 가능한 가치요인들이며, 각각의 하위 요인들에 대한 뉴스 기사의 가치 평가는 해당 뉴스 기사가 각각의 하위 요인들에 대해 얼마나 부합하는지를 판별하는 작업이다. 부합성에 대한 판별은 크게 규칙에 의한 판별과 통계적 분류에 의한 학습 등의 결과로 반환된다. 이때, 각각의 부합성 판별을 위한 기준을 본 연구에서는 적격질의(Competency Question, 이하 CQ)라고 명명하였다.

예를 들어 Table 2 의 하위 가치요인 중에 하나인 오피니언 리더 영향성에 대한 적격질의의 경우는 해당 기사에 대한 관찰 요소들인 뉴스의 메타데이터와 본문 내용에서 오피니언 리더의 이름이 얼마나 거론되는지를 검사하여 뉴스 기사에 대한 점수를 부여한다. 구체적으로는 각각의 뉴스 기사에서 사람들의 이름을 추출한 뒤, 각각의 추출된 이름이 오피니언 리더 사전에 등록된 이름에 해당하면 점수를 부과하는 방식이다.

Table 2. An example CQ algorithm for checking Opinion Leader Impact LLF

```

Example Algorithm : CQ algorithm for checking Opinion Leader Impact LLF
Input : newsArticle, opinionLeaderDic
Output : LLF Score

peopleNames <- extractEntityName(newsArticle, typePeople)
opinionLeaderDic <- loadDictionary("OpinionLeader")

for each aName in peopleNames (
    if aName is in opinionLeaderDic then
        observedOpinionLeaderNames.addName(aName)
    end if
end for

score <- calculateScore(observedOpinionLeaderNames)

Return score
    
```

이처럼 하위 가치요인의 평가에서는 각각의 하위 가치요인에 대한 고유의 적격질을 기준으로 뉴스 기사의 부합성을 검사하고 점수를 부여하게 된다. 다음 공식 (1)은 특정 뉴스기사 $k(\text{article}_k)$ 에 대해서 j 번째 하위 가치요인에 대한 점수인 $LLFValue_{jk}$ 를 계산 하는 공식이다. 이때 $CQCheck_k$ 는 검사 대상 뉴스기사에 대해 검사 기준이 되는 j 번째 LLF를 위한 CQ를 기준으로 부합성을 검사하여 검사결과를 반환하는 함수이다.

$$LLFValue_{jk} = CQCheck(\text{article}_k, CQ_j) \quad (1)$$

이와 같이 하위 가치요인 평가는 각각의 뉴스 기사를 분석해서 식별할 수 있는 정보요소들을 중심을 특정 조건인 CQ에 해당하는지 여부를 결정하고 이에 대한 점수를 부과한다. 이렇게 부과된 하위 가치요인 점수인 $LLFValue$ 는 이후 차상위 가치요인 평가의 중요한 기준이 된다.

3.3 Step V3: 차상위 가치요인(SLF) 평가

차상위 가치요인(Second Level Factor, SLF)에 대한 평가는 앞서 평가된 하위 가치요인의 점수를 이용하여 평가된다. 이때, 하나의 차상위 가치요인은 여러 개의 하위 가치요인으로 구성되며, 이때 각각의 하위 가치요인들은 중요도에 따라 가중치를 가지게 된다. 이때 가중치는 기본적으로는 각각의 하위 가치요인들 간에 균일한 값을 가지지만, 가치요인의 최적화 과정에서 조정될 수 있다. 다음 공식 (2)는 차상위 가치요인을 평가하는 공식이다.

$$SLFValue_{ik} = \sum_{j=0}^n (W_{ij} \times LLFValue_{ijk}) \quad (2)$$

$$= \sum_{j=0}^n (W_{ij} \times CQCheck(\text{article}_k, CQ_{ij}))$$

$$W_{ij} = \text{weight of } LLF_{ij}, \sum W_{ij} = 1$$

공식 (2)에서는 뉴스기사 k 에 대한 i 번째 SLF에 대한 차상위 가치요인의 점수인 $SLFValue_{ik}$ 를 계산하는 방법을 보여주고 있다. 각각의 하위 가치요인 점수에 LLF별 가중치를 반영하는 방식으로 가치를 평가한다. 이때 W_{ij} 는 i 번째 SLF를 구성하는 j 번째 LLF에 대한 가중치 W 로서, 하나의 SLF에 대한 LLF 가중치의 총합은 1이다. 이렇게 계산된 차상위 가치요인 점수 $SLFValue$ 는 최상위 가치요인 평가를 위한 기준이 된다.

3.4 Step V4: 최상위 가치요인(TLF) 평가

뉴스 기사에 대한 최종적인 가치 판단의 기준은 최상위 가치요인(Top Level Factor)을 기준으로 한다. 이때 최상위 가치요인의 평가는 앞서 계산된 차상위 가치요인 점수인 $SLFValue$ 에 Table 1의 차상위 가치요인별 가중치(Second Level Factor Weight)를 반영하여 계산한다. 다음 공식 (3)은 최상위 가치요인을 평가하는 공식이다.

$$TLFValue = \sum_{i=0}^m \alpha_i \times SLFValue_{ik} \quad (3)$$

$$= \sum_{i=0}^m \alpha_i \times \sum_{j=0}^m W_{ij} \times LLFValue_{ijk}$$

$$= \sum_{i=0}^m \alpha_i \times \sum_{j=0}^m W_{ij} \times CQCheck(\text{article}_k, CQ_{ij})$$

$$\alpha_i = \text{weight of } SLF_i, 0 \leq \alpha_i \leq 1,$$

$$W_{ij} = \text{weight of } LLFValue_{ij}, \sum W_{ij} = 1$$

공식 (3)에서는 뉴스 기사 k 에 대한 최상위 가치요인 점수를 계산하며, 각각의 차상위 가치요인 점수에 반영되는 가중치인 α_i 는 신문 편집자 혹은 뉴스 소비자들의 의견에 의해 평가되어 결정되는 가중치로 0~1사이의 값을 갖는다. 본 연구에서 적용한 Table 1의 값은 초기 값으로 적용하였으며 추천의 품질을 최적화하기 위해 지속적으로 조정될 수 있다.

4. 온라인 뉴스 추천 개인화 절차 및 기법

3장의 뉴스 가치평가의 결과인 뉴스 기사 순위점수(Ranking-Scores)는 뉴스 소비자 개인의 선호가 반영되지 않은 상태이다. 4장에서는 뉴스 소비자의 특성을 고려해서 뉴스 기사 순위점수를 개인화한다.

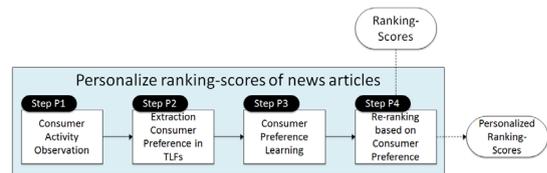


Fig. 3. Personalization Process for News Ranking Scores based on VNP

전체 개인화의 절차는 Fig. 3과 같이 Step P1. 소비자

행위 관찰(Consumer Activity Observation), Step P2 Low Level Factor Valuation), Step P2. TLF로부터의 소비자 선호 추출(Extraction Consumer Preference in TLFs), Step P3. 소비자 선호 학습(Consumer Preference Learning), Step P4. 소비자 선호 기반 랭킹 조정 (Re-ranking based on Consumer Preference) 등 4단계의 절차로 진행된다.

Step P1. 소비자 행위 관찰 단계에서는 온라인 뉴스 소비 과정에서 관찰될 수 다양한 뉴스 소비 활동을 관찰한다. 기사 목록을 조회하고(list), 기사의 내용을 읽고(read), '좋아요' 혹은 별점을 부여하여 자신의 감정을 표현하고(sentiment feedback), 댓글 혹은 답글을 달거나 의견을 개진하는(comment) 등의 활동에 대한 이력을 기록하고 관찰한다.

Step P2. TLF로부터의 소비자 선호 추출 단계에서는 관찰된 소비 활동들을 분석하여 뉴스 소비자 개인의 선호를 추출한다. 이때 추출되는 선호 특성은 뉴스 기사의 키워드나 주제어 혹은 주제 분류에 대한 선호가 아닌, 앞서 제시한 최상위 가치요인(TLF)들에 대한 선호를 나타낸다.

Step P3. 소비자 선호 학습 단계에서는 추출된 선호 특성을 지속적으로 학습하여 선호의 최신성을 유지하는 단계이다. 소비자 개인의 선호 특성은 소비 활동을 관찰하여 분석하는 과정에서 관찰되는 동안 계속해서 변화하게 된다. 그러므로 소비자 개인의 선호를 제대로 반영하기 위해서는 선호 특성의 변화를 지속적으로 학습하는 과정이 필요하다.

Step P4. 소비자 선호 기반 랭킹 조정 단계에서는 학습된 개인 선호를 이용하여 3장의 결과물로 제공되는 뉴스 기사들에 대한 순위점수를 재조정한다. 재조정된 뉴스 기사들의 순위점수는 이후 온라인 뉴스 서비스의 중요 뉴스를 결정하는 개인화된 기준으로 이용된다.

4.1 Step P1: 소비자 행위 관찰

소비자 행위 관찰 단계에서는 앞서 Step V1. 정보 관찰단계에서 획득된 정보들 중 뉴스 소비자 행위에 대한 정보를 이용한다. 뉴스 소비자의 행위는 뉴스 서비스가 제공되는 과정에서 관찰되는 사용자 행위를 의미적으로 분류한 것으로 서비스 사용기록을 제공받을 수 있는 경우에는 서비스 사용 기록으로부터 추출한다. 반면에 타 사업자의 뉴스 서비스에 나타난 행위정보의 경우에도 웹

크롤링 기술을 이용해서 수집할 수 있다.

Table 3. Observation of Consumer Activities in ZinyNews

Observable Consumer Activity	Description	Initial Activity Weight
List View	View a article list	N/A
Read Article	Read a article	1.0
SFeedback	Express a sentimental feedback(ex. like, dislike, star, love etc.)	1.2
Comment	Write some comments for each article	1.5
Share	Share a article with other consumer via SMS, e-mail, SNS, blog etc.	1.8
Scrap	Keep a article for reading after	2.0

본 연구에서 사용자 행위는 Table 3 와 같이 모바일 기반 온라인 뉴스 서비스인 지니뉴스에서 관찰 가능한 행위들을 중심으로 리스트 보기(List View), 기사 읽기(Read Article), 감성 피드백(SFeedback), 댓글(Comment), 공유(Share), 스크랩(Scrap) 등으로 분류하였다. 각각의 행위들에 대해 사용자 선호를 나타내는 행동 점수로서 행위 가중치(Activity Weight)를 설정한다. 행위 가중치는 각각의 행위들이 해당 기사에 대한 관심의 정도를 나타내는 점수로서 실험을 통해 조정될 수 있는 값이다.

본 연구에서는 각각의 관찰되는 행동들에 대한 초기 행위 가중치는 온라인 뉴스 소비 과정에서 예상되는 소비자 관여의 정도를 기준으로 설정하였다. 우선 리스트 보기의 경우는 소비자의 관여가 거의 없으므로 0으로 초기 값을 설정하였다. 그리고 기사 읽기의 경우는 기본적인 사용자 관심의 크기로 판단하여 1로 설정하였다. 또한, 사용자 관심의 정도가 가장 크다고 판단되는 행위인 스크랩의 경우는 2로 초기 값을 설정하였다. 감성 피드백과 댓글, 공유의 경우는 기사 읽기와 스크랩의 행위 가중치를 기준으로 사용자 개입의 정도를 각각의 시스템의 기능적 복잡도에 따라 1.2, 1.5, 1.8로 초기 설정하였다. 이렇게 초기 설정된 가중치는 실제 서비스를 제공하는 과정에서 경험적으로 조정될 수 있다고 가정하였다.

4.2 Step P2: 최상위 가치요인에 대한 소비자 선호 추출

본 연구에서 추출하고자 하는 뉴스 소비자의 선호는 뉴스 기사에 대한 주제 분류나 주제어를 중심으로 하는

내용 기반 선호가 아닌, 뉴스 기사의 뉴스 가치에 기반을 둔 선호이다. 이때 뉴스 가치는 앞서 3장에서 정의된 뉴스 가치모델 상의 최상위 가치요인(Top Level Factor, TLF)들을 의미하며, 이러한 이유로 가치기반 뉴스 선호(Value based News Preference, VNP)라고 정의하였다.

즉, 가치기반 뉴스 선호는 소비자가 선호를 나타내는 뉴스 기사들을 이용하여 뉴스 가치모델 상의 최상위 가치요인에 대한 각 소비자들의 선호 특성을 추출한 것이다.

자세한 VNP 추출 과정은 다음과 같다.

- 1) 우선 정해진 기간 동안의 행위 이력들로부터 사용자 선호를 반영한 행위들을 선별한다.
- 2) 각각의 행위들을 분류하고 각 행위 분류에 대한 행위가중치를 부여한다.
- 3) 3장에서 평가된 행위 대상 기사의 TLF 값에 관찰된 행위가중치를 반영한다.
- 4) 각각의 TLF별로 행위가중치가 반영된 TLF값의 평균을 계산한다.

이러한 과정을 수행하는 계산 공식은 공식 (4)와 같다.

$$VNP_{ij} = \frac{\sum_{k=0}^m TLFValue_j \times ActivityWeight_{ijk}}{m} \quad (4)$$

$i = consumer, j = TLF, k = observedactivity$

공식 (4)에서는 뉴스 소비자 i의 뉴스 기사 선호 특성을 각각의 최상위 가치요인 j 별로 계산하는 공식이다. 여기서 뉴스 소비자의 관찰된 행위 k는 각각의 행위의 종류에 따라 각 행위 별로 선호 가중치인 $ActivityWeight_k$ 를 부여 받게 된다. 그리고 여기에 각각의 행위의 대상이 되는 뉴스기사들의 최상위 가치요인 점수 $TLFValue_j$ 를 반영하면 각각의 행위 k로 관찰되는 뉴스 소비자의 특성이 된다. 최종적으로는 이러한 값들의 평균을 내어 특정 행위 이력들로부터 관찰된 뉴스 소비자의 뉴스 가치 기반 선호특성을 추출하게 된다.

Fig. 4 는 TLF 값에 행위가중치를 반영하는 과정에 대한 예시이다. 읽기(read), 평가(sfeedback), 댓글(comment) 등의 뉴스 기사 소비 행위에 대해, 행위가중치와 대상 뉴스 기사의 TLF 가치를 이용하여 각각의 소비 활동에 대한 가중치가 반영된 TLF 값(Weighted TLF Value)을 계산한다.

Consumer Activity	
Date	2014-11-22
Consumer ID	User0020
Activity Type	sFeedback
Activity Weight	1.20
News ID	#1034

TLF Value		Weighted TLF Value	
Social Significance	65%	Social Significance	78%
Visual Novelty	56%	Visual Novelty	67%
Audience Relevance	28%	Audience Relevance	34%
Human Interest	5%	Human Interest	6%

X 1.20

Fig. 4. Example of Weighted TLF Value Calculation

이렇게 가중치가 반영된 TLF 값에 대해 각 기간별 평균값을 계산하여 최종적인 가치기반 선호 특성(VNP)을 추출한다. 선호를 추출하는 과정은 매우 거대한 연산을 요구하므로, 특정 기간 단위로 선호를 계산하여 취합하는 것이 좋다. 만약 전체 기간에 대한 선호를 구하는 경우는 각 기간별 선호와 각 기간에 대한 행위 규모를 이용하여 가중평균을 계산하면 된다.

Fig. 5 는 VNP 추출이 한 예로서, 앞서 계산된 TLF 평균값에 각 기간별 행위 가중치를 반영한 후, 이에 대한 평균값을 이용하여 해당 뉴스 소비자에 대한 VNP를 추출한다.

Observed VNP _{Period#1}		Observed VNP _{Period#n}	
Social Significance	37%	Social Significance	19%
Visual Novelty	17%	Visual Novelty	20%
Audience Relevance	28%	Audience Relevance	52%
Human Interest	18%	Human Interest	9%
Activity Scale	217	Activity Scale	413
Activity Weight	117%	Activity Weight	22%

Weighted Average VNP _{Period#1~n}	
Social Significance	33%
Visual Novelty	21%
Audience Relevance	41%
Human Interest	15%

Fig. 5. Example of VNP Calculation

정리하자면 VNP는 특정 시점에서 특정 뉴스 소비자가 어떤 뉴스 가치요인에 선호를 가지고 있는지를 나타

내게 된다. 이때, 뉴스 서비스를 처음 사용하는 뉴스 소비자의 경우에는 기록된 활동이 없거나 매우 적으므로 VNP에 대한 초기 값(Default)을 부여한다. 본 연구에서는 초기 값 모두 균등하게 25%로 설정했으나, 실제 서비스 운영에서는 이미 많은 뉴스 소비자들이 뉴스 서비스를 사용하고 있으므로, 새로운 뉴스 소비자가 처음으로 서비스를 이용하는 시점에서 전체 사용자의 VNP 평균을 적용하는 것이 타당하다. 이렇게 해서 추출된 VNP를 2차원 영역에 시각화하면 Fig. 6.과 같이 표현할 수 있다.

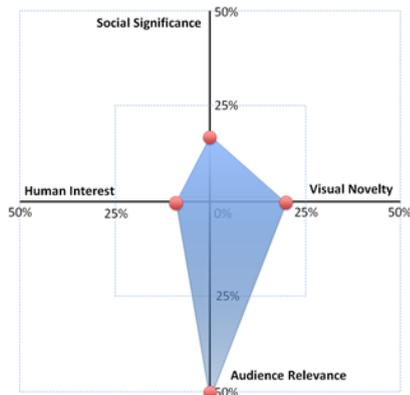


Fig. 6. VNP Visualization

또한, 본 연구에서는 VNP를 구성하는 각각의 최상위 가치요인을 사회적 중요도(social significance), 새로운 볼거리(visual novelty), 수용자 관련성(audience relevance), 인간적 흥미(human interest) 등 4개의 TLF를 기준으로 하였지만, 각각의 최상위 가치요인의 구성은 확장 가능하다.

4.3 Step P3: 소비자 선호 학습

뉴스 소비자가 뉴스 서비스를 이용하는 과정에서 기록되는 행위 정보에 따라 VNP는 지속적으로 변화하게 된다. 또한, 시스템 구성에 의해 뉴스 소비자는 자신의 선호를 직접 지정할 수도 있다. 그러므로 시스템은 이러한 사항들을 모두 반영하여 VNP를 학습할 수 있어야 한다.

Table 4의 알고리즘처럼 VNP의 학습은 총 4가지 경우로 나뉘어 진행된다.

- 1) 뉴스 소비자가 새로운 사용자이고(추출된 VNP

(extractedVNP)가 없는 경우), 뉴스 소비자가 지정한 VNP가 없는 경우(consumerDefinedVNP)에는 초기 VNP(defaultVNP)를 적용한다.

- 2) 뉴스 소비자의 이력을 통해 VNP를 추출할 수 있으나 뉴스 소비자가 지정한 VNP가 없는 경우는 추출된 VNP를 적용한다.
- 3) 추출된 VNP가 없고, 뉴스 소비자가 지정한 VNP가 있는 경우에는 지정된 VNP를 적용한다.
- 4) 추출된 VNP가 있는 상태에서, 뉴스 소비자가 지정한 VNP가 있는 경우에는 추출된 VNP와 지정된 VNP를 모두 반영하되 학습 가중치(learningWeight) 만큼 추출된 VNP를 반영한다. 이때, learningWeight는 0부터 1사이의 값을 갖는 시스템 조정 값이다.

Table 4. Learning VNP

Algorithm : Learning VNP
Input : extractedVNP, consumerDefinedVNP, learningWeight
Output : learnedVNP
<pre> If (extractedVNP == nil and consumerDefinedVNP == nil) then learnedVNP <- defaultVNP Else If (extractedVNP != nil and consumerDefinedVNP == nil) then learnedVNP <- extractedVNP Else If (extractedVNP == nil and consumerDefinedVNP != nil) then learnedVNP <- consumerDefinedVNP Else learnedVNP <- learningWeight * extractedVNP + (1- learningWeight) * consumerDefinedVNP End if Return learnedVNP </pre>

4.4 Step P4: 소비자 선호 기반 랭킹 조정

추출된 커버스토리 기사 목록을 개인화하기 위해서는 개인화 되지 않은 커버스토리 기사 목록에서의 순위를 뉴스 소비자의 개인 선호에 따라 조정해야 한다. 이때 추출된 VNP는 뉴스 소비자의 TLF에 대한 개인 선호를 나타내며, 또한 각각의 뉴스 기사들은 측정된 TLF별 뉴스 가치에 따라 색인되어 있다. 그러므로 뉴스 기사 순위에 대한 재조정은 뉴스 소비자의 VNP에 나타난 각 TLF에

대한 선호를 해당 뉴스의 TLF 가치에 가중치로 반영함으로써 이루어진다.

뉴스 기사의 순위 재조정을 아래 공식 (5)와 같다. 이때 i 는 TLF의 번호이다.

$$rerankedScore_{kc} = \sum_{i=0}^n (TLFWeight_{ci} \times TLFValue_{ki}) \tag{5}$$

공식 (5)에서 뉴스 소비자 c 의 VNP를 반영한 뉴스 기사 k 에 대한 조정된 순위값인 $rerankedScore_{kc}$ 은 해당 뉴스 기사 각각의 TLF 가치 $TLFValue_{ki}$ 에 대해 VNP에 표현된 TLF 선호의 가중치인 $TLFWeight_{ci}$ 를 반영한 후 전체 값들을 합산하여 계산할 수 있다.

5. 실험

본 연구의 실험에서는 1면기사(커버스토리)로 선정된 뉴스 기사들에 대한 실제 뉴스 소비자들의 만족도를 기준으로, 본 연구에서 제시한 방법이 유효한 지를 검증하였다.

이를 위해, 실험 대상 뉴스 기사는 국내 주요 온라인 뉴스 서비스들인 N사, D사, G사, C사에서 제공하는 동일 시간대의 중요 기사들 중 상위 10개 기사가 선정되었다. 다만 G사의 경우 서비스 특성상 상위 7개만의 뉴스 기사들을 제공하여 7개만을 실험대상으로 선정하였다.

또한, 실험 대상 뉴스 소비자는 온라인 뉴스 사용의 주 연령대인 20대부터 50대 사이의 남성과 여성 소비자들을 대상으로 하였으며, 전체 조사 대상 중 응답을 한 21명의 뉴스 소비자를 실험 대상으로 선정하였다.

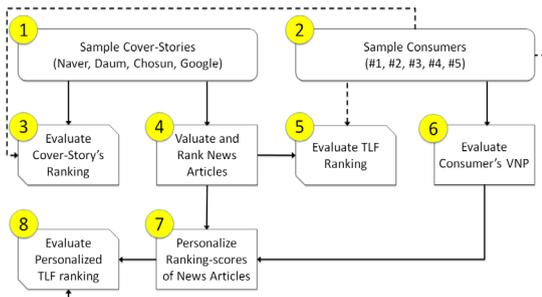


Fig. 7. The process of experiment

본 연구의 실험 절차는 Fig. 7.과 같다.

- ① 우선 온라인 뉴스 서비스를 제공하는 N사, D사, C사, G사의 특정 일시의 중요기사를 실험대상 기사로 선정하였다.
- ② 조사대상 소비자 중 뉴스 기사에 대한 선호특성의 차이가 확인한 5명의 소비자를 실험대상으로 선정하였다.
- ③ 각 온라인 뉴스 서비스의 중요기사들에 대한 실험대상 소비자들의 만족도를 평가하였다.
- ④ 각 온라인 뉴스 서비스의 중요기사들에 대해 하위 가치요인 평가, 차상위 가치요인 평가, 최상위 가치요인 평가를 수행하여 통합 순위점수(TLF Ranking)를 계산하였다.
- ⑤ 통합 순위점수를 기준으로 선정된 뉴스 기사들에 대해 실험대상 소비자들의 만족도를 평가하였다.
- ⑥ 실험대상 소비자들의 가치 기반 뉴스 선호인 VNP 특성을 계산하였다.
- ⑦ 조사된 각 소비자들의 VNP 특성을 이용하여 통합 순위점수를 개인화하였다.
- ⑧ 개인화된 통합 순위점수에 대해 실험대상 소비자들의 만족도를 평가하였다.

이때, 만족도는 공식 (6)의 정확도(precision)와 재현율(recall)에 근거한 F1-score를 이용하여 측정하였다.

$$F1 = 2 \times \frac{precision \cdot recall}{precision + recall} \tag{6}$$

여기서 정확도는 추천되어 제시된 결과 항목들 중에서 소비자가 원하는 결과 항목들의 비율이며, 재현율은 제시되어야 하는 결과 항목들에 대한 제시된 결과 항목들의 비율을 의미한다[11].

5.1 실험대상 하위 가치요인 선정

실험 환경을 구성하는 데에 있어서 가장 중요한 부분 중에 하나는 3.2절에 제시된 하위 가치요인 중 평가 가능한 하위 가치요인을 선정하는 것이다. 이때, 하위 가치요인은 실험환경에서 사용 가능한 분석기술을 통해 측정되게 된다. 그러므로 해당 분석 기술로 적격질의(CQ)의 구성이 가능한 하위 가치요인들을 Table 5과 같이 선정하였다.

우선 오피니언 리더 영향성(Opinion Leader Impact)

에 대해서는 개체명 추출기(Entity Name Extractor)를 이용하여 뉴스 기사 내용으로부터 통해 인명을 추출하고, 해당 인명이 오피니언 리더 사전에 존재하는 지를 기준으로 부합성을 평가한다. 개체명 추출기는 주어진 텍스트에서 인물, 기관, 건물, 제품, 지명, 사건 등 알려진 개체명들을 자동으로 추출해 내는 텍스트 분석 분야의 기술이며, 사회적인 오피니언 리더 사전(Social Opinion Leader Dictionary)는 사회적으로 중요한 인물들의 명단에 대한 개체명 사전으로 개체명 추출기는 해당 사전을 이용하여 작동된다.

정치적 중요성(Political Importance)에 대해서는 주제어 특성 추출기(Topic Feature Extractor)를 이용하여 뉴스 기사 내용에서 부각되는 중요 주제어들을 선별하고, 해당 주제어가 정치 주제 사전(Political Topic Dictionary)에 존재하는지를 기준으로 부합성을 평가한다. 주제어 특성 추출기는 주어진 텍스트에서 키워드들을 추출하고, 추출된 키워드 중 의미적으로 중요한 키워드나 키워드 군을 주제로 선정하여 추출하는 기술이다.

Table 5. Experimental Low Level Factor

First Level Factor	Second Level Factor	Experimental Low Level Factor	System Component
Social Significance	Impact	Opinion Leader Impact	Entity Name Extractor, Social Opinion Leader Dictionary
	Prominence	Political Importance	Topic Feature Extractor, Political Topic Dictionary
	Depth	N / A	N / A
	Conflict	Structure of Conflict	Topic Feature Extractor, Conflictive Issue Dictionary
Visual Novelty	Novelty	N / A	N / A
	Action	Visual Content	Qualifying Rule, Multimedia Resource Metadata
Audience Relevance	Usefulness	Statistics	Keyword Matching, Useful Component Expression Dictionary
	Proximity	N / A	N / A
Human Interest	Entertainment	Entertainingness	Classification, Category Set
	Story	N / A	N / A

대립 구조(Structure of Conflict)에 대해서는 정치적 중요성과 마찬가지로 방법을 이용하여 뉴스 기사 내용으로

부터 중요 주제어를 선별하고, 해당 주제어가 대립적 이슈 사전(Conflictive Issue Dictionary) 사전에 등록되어 있는지를 기준으로 부합성을 평가한다.

시각적 내용(Visual Content)에 대해서는 뉴스 기사에 포함된 멀티미디어 정보의 메타데이터(Multimedia Resource Metadata)를 이용해서 시각적으로 의미 있는 멀티미디어 정보가 포함되었는지에 대한 판별 규칙(Qualifying Rule)을 적용하여 부합성을 판단한다. 이때 의미 있는 멀티미디어 정보란 기사에 포함되어 연결된 리소스들 중 이미지, 동영상, 소리 등과 같은 멀티미디어 리소스가 존재하는 지와, 존재한다면 충분한 규모(이미지의 크기나 해상도, 동영상의 길이나 크기, 소리의 길이 등)가 되는 리소스인지를 판별하는 규칙이다.

통계(Statistics)에 대해서는 뉴스 기사에 포함된 키워드 중에서 특정 통계를 나타내는 표나 그림을 표현하는 키워드가 있는 지를 기준으로 부합성을 판단한다. 이때 뉴스 기사의 본문 내용뿐만 아니라, 뉴스 기사를 기술한 HTML상의 태그 상에도 관련된 표현이 있을 수 있으므로, 기사 내용과 함께 분석대상이 된다.

오락성(Entertainingness)에 대해서는 뉴스 기사를 주제 분류 체계(Category Set)에 따라 분류(Classification)한 뒤 해당 기사가 오락(Entertainment)과 관련된 분류에 해당되는 지를 기준으로 부합성을 평가한다. 이때 분류는 텍스트 분석 기술 중 자동 분류 기술을 이용한다.

실험 시스템의 구성은 Fig. 8과 같이 구성하였다. 정보 수집, 색인 및 검색, 언어 분석 및 개체명 추출기, 콘텐츠 분류기, 토픽 추출기에 대해서는 각각 상용 제품인 [IN2] Tornado, [IN2] DOR, [IN2] LEA, [IN2] HBC, [IN2] TopicRank를 이용하였고, 관련하여 개체명 사전, 분류 모델, 토픽 사전 등은 상용 서비스인 지니뉴스에서 제공한 사전과 분류모델을 적용하였다[12].

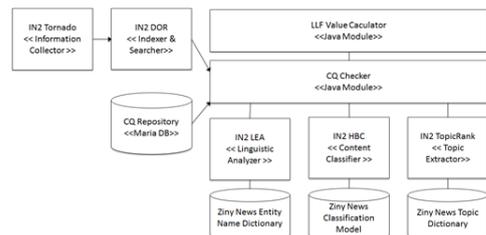


Fig. 8. Experimental System Architecture

이때, 오피니언 리더 영향성, 정치적 중요성, 대립 구

조 등에 대한 부합성 평가 과정에서 사용되는 오피니언 리더 사진, 정치 주제 사진, 대립적 이슈 사진은 개체명 식별 기술(IN2 LEA)과 토픽 추출 기술(IN2 TopicRank)은 통해 식별된 개체명과 토픽들에 대해서 사진 편집자인 큐레이터들을 통해 선별된 사진을 이용하였다.

5.2 실험 결과 및 평가

1) 실험 대상 뉴스 기사의 선정과 순위점수 계산

우선 실험을 위해 N사, D사, G사, C사 등 국내 주요 온라인 뉴스 서비스에서 제공하는 특정시점(2015년 3월 29일 22시 기준)의 중요 기사들을 선정하였고, 선정된 뉴스 기사들에 대한 하위요인 가치를 측정하였다. 하위 가치요인 평가에 있어서는 5.1에서 선정한 평가 가능한 하위요인 가치들을 대상으로 하였으며, 평가된 하위 가치요인 가치들을 기반으로 차상위 가치요인과 최상위 가치요인을 평가하고 해당 뉴스들의 순위점수(TLF Value)를 계산하였다 [Fig. 9].

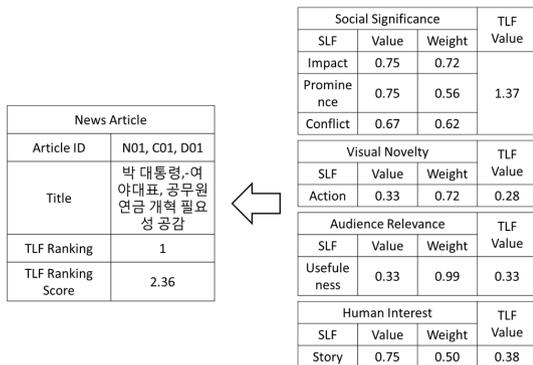


Fig. 9. Ranking Result base on TLF Value

Fig. 9에서는 뉴스 기사 N01(다른 언론사 뉴스인 C01과 D01과 같은 내용의 기사임)에 대해서 사회적 중요도, 새로운 볼거리, 수용자 관련성, 인간적 흥미 등의 4 가지 최상위 가치요인에 대한 가치를 계산하고, 이를 기반으로 중요 순위를 결정한다. 이 때, 각각의 최상위 가치요인들도 가중치를 갖게 되지만 이는 시스템 운영시의 조정 값이다. 본 연구에서는 모든 최상위 가치요인이 동등한 가중치를 갖는다고 전제하였다.

2) 실험 대상 소비자 선정

실험 대상 소비자는 연령과 성별을 고려하지 않고 무

작위로 선정된 소비자 21명을 대상으로 하였다.

또한, 개인화된 순위점수를 평가하기 위해, 전체 실험 대상 소비자 후보군 중 뉴스 가치 기반 선호(VNP) 특성이 상대적으로 차이가 뚜렷한 5명의 실험대상 소비자를 선정하였다. 이때, 소비자의 선호 특성은 뉴스를 소비한 이력을 중심으로 소비자들이 실제로 소비한 뉴스들의 특성을 중심으로 선호 특성을 추출하여야 한다. 다만 본 연구에서는 비교 대상이 되는 뉴스 서비스들의 실질적인 뉴스 소비 이력을 평가할 수가 없으므로, 각각의 실험대상 소비자 후보군들에 대해 가치기반 뉴스 선호에 대한 각각의 소비자들의 설문 결과를 기준으로 판단하였다.

Fig. 10은 개인화된 순위점수를 평가하기 위해 선별된 5명의 실험대상 소비자들의 뉴스 가치기반 선호에 대한 차이를 보여준다. 선정된 실험 대상 소비자들의 선호 특성에 따르면 최상위 가치요인인 사회적 중요도, 새로운 볼거리, 수용자 관련성, 인간적 흥미 등에 대해서, 사회적 중요도와 수용자 관련성의 경우는 공통적으로 높은 뉴스 기사 선호를 나타내었으며, 인간적 흥미와 새로운 볼거리의 경우는 소비자의 성향에 따라 매우 상이한 뉴스 기사 선호를 나타내고 있다.

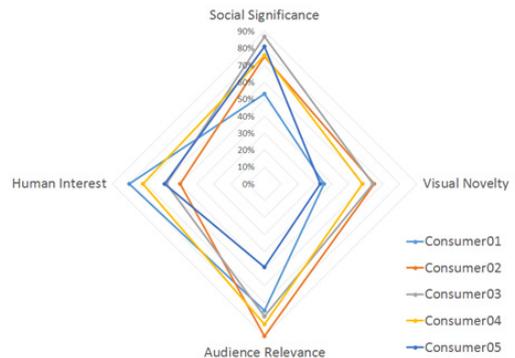


Fig. 10. The comparison of value based news preferences

3) 순위 점수의 개인화

계산된 순위점수를 개인화하기 위해서 Table 6]와 같이 각각의 TLF 값에 대해 각 사용자들이 가지고 있는 TLF에 대한 선호 특성인 가중치(W)를 적용하여 개인화된 순위점수(RS)를 계산하였다.

Table 6. Personalized Ranking Score for Consumer01

AID	Consumer01								RS
	SS		VN		AR		HI		
	V	W	V	W	V	W	V	W	
N01	1.37	0.53	0.28	0.35	0.33	0.75	0.38	0.80	1.37
N02	0.62	0.53	0.55	0.35	0.66	0.75	0.13	0.80	1.12
N03	0.41	0.53	0.28	0.35	0.66	0.75	0.13	0.80	0.81

...
 AID = News Article ID, SS = Social Significance, VN = Visual Novelty, AR = Audience Relevance, HI = Human Interest, RS = Ranking Score, V = Value, W = Weight

개인화된 순위점수(Personalized Ranking Score)를 기준으로 각각의 소비자를 위한 중요 뉴스를 선정하면 Table 7과 같이 특정 시점에서의 중요 기사 추천 목록이 각각의 소비자의 선호 특성에 따라 달라짐을 확인할 수 있다.

Table 7. Personalized Recommendation Result

Ranking	TLF		Personalized TLF (Consumer01)	
	ID	Title	ID	Title
1	N01	Presid..	N01	Presid..
2	D06	4 billio...	G06	Withou..
3	G06	Withou..	D06	4 billio...
4	N02	Young...	D04	Japan..
5	N05	Wing...	N05	Wing...
6	N06	Posco..	N02	Young...
7	D04	Japan..	N10	Insura..
8	N10	Insura..	D07	Asbest..
9	C03	Conser..	N06	Posco..
10	D07	Asbest..	G05	Hyund..

4) 추천결과의 평가

추천 결과의 평가를 위해 제공된 실험 대상 뉴스 기사들 중 각각의 실험 대상 소비자들이 선택한 중요 뉴스 기사들을 정답 셋으로 선정하였다. 그리고 이렇게 선정된 정답 셋과 N사, D사, G사, C사 및 본 연구의 순위점수를 기준으로 선정된 중요 뉴스 기사 목록을 비교하여 정확도와 재현율을 측정하고 이를 토대로 F1 점수로 환산하여 만족도를 산출하였다.

Table 8. Evaluation of Recommendation (n=21, unit:%)

	N	D	G	C	TLF
Average Precision	26.00	50.00	20.00	38.00	56.00
Average Recall	27.22	52.22	15.00	40.00	58.89
Average F1 Score	26.51	50.91	17.11	38.85	57.22

Table 8의 평가 결과를 살펴보면 본 연구에서 제시한 순위점수 기반의 추천의 경우 추천의 만족도(21명 실험 대상 소비자 군의 평균 F1 Score)는 57.22%로 비교 대상 군인 각 온라인 뉴스 서비스들과 비교하였을 때 상대적으로 우수함을 확인할 수 있다.

또한, 개인화된 순위점수의 효과를 측정하기 위해 앞서 선택된 5명의 비교군의 가치기반 뉴스 선호 특성을 반영한 결과는 Table 9과 같이 정확도의 경우 17.86%, 재현율의 경우 17.92%로 각각 상승하였으며, 전체적인 추천의 만족도(F1 Score)의 경우도 16.79% 상승한 것으로 관찰되었다.

Table 9. Comparison of TLF and PTLF (unit:%)

	TLF	PTLF	Improvement
Average Precision	56.00	66.00	17.86
Average Recall	58.89	69.44	17.92
Average F1 Score	57.22	66.83	16.79

6. 결론

본 연구에서는 전통적인 지면 신문에서 중요 뉴스(커버스토리)를 선정하는 방법을 온라인 뉴스 서비스에 적용하는 방법을 제안하였다. 이와 함께 뉴스 가치 기반의 중요 뉴스 선정 기준에 따른 개인의 선호 특성과 이를 이용한 개인화 방법을 제안하였다. 본 연구에서 적용한 뉴스 가치 기반 선호는 뉴스 기사의 내용적인 특성을 중심으로 개인 선호만이 아닌, 뉴스의 소비자(수용자) 개인과 개인이 속해있는 사회 환경에 대한 영향성과 중요도, 뉴스 소비자들의 새로운 소비 욕구를 느끼게 할 만한 내용적 특성들과 뉴스 내용 구성 자체의 흥미성 등이 포함된 다양한 관점의 '뉴스 기사 가치 구조 모델(News Value Structure Model)'을 근간으로 하고 있다.

또한, 이러한 모델을 이용하여 중요 뉴스를 순위 점수를 측정하기 위한 절차적 프레임워크와 주요 기법들을 제시하였고, 뉴스 가치 기반의 중요 뉴스 선정이 더 효과적임을 실험실 환경 내에서 확인하였다. 이와 더불어 뉴스 가치 구조 모델에 따른 가치기반 뉴스 선호에 따라 추천 결과를 개인화함으로써 추천 결과의 만족도가 더욱 향상됨을 확인하였다.

본 연구에서 제안한 뉴스 가치 구조 모델을 적용하는

방법을 통해 통상적인 콘텐츠 추천의 취약성을 보완한 중요 뉴스 선정이 가능하다. 이렇게 함으로써 뉴스 소비자들로 하여금 원하는 뉴스를 좀 더 빠른 시간에 발견할 수 있도록 도울 수 있다. 특히 이러한 향상은 모바일이 중심이 된 콘텐츠의 소비 환경에 더욱 적합하다.

후속 연구로는 제시된 모델과 기법을 실제 운영 시스템에 적용하여 대규모 운영환경에서의 품질 측정과 품질 향상에 대한 연구를 진행하려 한다.

References

[1] S. S. Kim, "A study of news production strategy for consumer experiential value", Korean Journal of Journalism & Communication Studies, Volume 57, Issue 2, pp.33~37, 2013

[2] J. H. Lee, W. Y. Kim, S. M. Kang, Y. J. Choi, "A Comprehensive and Structural Approach to News Values in Multimedia Environment - Extraction of a Structural Model of News Values", Korean Journal of Broadcasting, Volume 27, Issue 1, pp.167-212, 2013

[3] C. B. Jung, T. H. Kim, H. C. Cheon, J. M. Choi, "A News Recommendation System based on Document Clustering Using WordNet", Journal of KIISE, Volume 35, Issue 2, pp.258-262, 2008

[4] Yonghui Wu, Yuxin Ding, Xiaolong Wang, Jun Xu, "Topic based automatic news recommendation using topic model and affinity propagation", Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, pp. 1299-1304, 2010

[5] J. Park, W. S. Hwang, S. W. Kim, "A Method for Recommending Newly Updated News", Proceedings of the Korea Computer Congress (KCC) 2013, pp. 289-290, 2013

[6] Husna Sarirah Husin, James A. Thom, Xiuzhen Zhang, "News recommendation based on Web usage and Web content mining", Data Engineering Workshops (ICDEW), 2013 IEEE 29th International Conference, 326 - 329, 2013

[7] Won-Jo Lee, Kyo-Joong Oh, Chae-Gyun Lim, and Ho-Jin Choi, "User Profile Extraction from Twitter for Personalized News Recommendation", Advanced Communication Technology (ICACT), 2014 16th International Conference on, 779-783, 2014

[8] T. Song, "Design and Implementation of Personalized News Recommendation System Considering User Reading Habit under Smartphone Environment", Journal of the Korea Institute of Information and Communication Engineering, Volume 18, Issue 7, pp. 1628-1633, 2014

[9] J. A. Yang, "News Story Salience and Users? Selective Exposure - Effects of Popularity Indications on Online News Exposure", Korean Journal of Broadcasting, Volume 25, Issue 2, pp. 77-117, 2011

[10] Y. D. Seo, J. H. Ahn, "Contents Recommendation

Method Based on Similarity of Users Viewing Tendency", Journal of Korean Institute of Information Technology, Volume 10, Issue 9, pp. 187-196, 2012

[11] J. M. Lee, J. Park, W. S. Hwang, S. W. Kim, "Accuracy Evaluation of New Recommendation Methods in Online News Services", Proceedings of KIISE 2013 Fall Conference, pp. 320-321, 2013

[12] Saltlux Inc., "Smart Data Products", <http://smartdata.saltlux.com/product/>. 2015

최 광 선(Kwang Sun Choi)

[정회원]



- 1995년 2월 : 인하대학교 자원공학 (학사)
- 1995년 3월 ~ 2000년 1월 : 대우 정보시스템 기술연구소 대리
- 2000년 2월 ~ 2002년 11월 : 에이 전트리더 기술연구소 실장
- 2003년 11월 ~ 2005년 3월 : 큐브 테크 기술연구소 차장
- 2005년 9월 : 숭실대학교 소프트웨어 공학 (석사)
- 2005년 6월 ~ 현재 : 솔트룩스 CDS 사업본부 본부장
- 2008년 8월 : 숭실대학교 컴퓨터학 (박사수료)

<관심분야>

서비스 지향 아키텍처(SOA), 모바일 서비스(Mobile Service), 객체지향 S/W공학, 컴포넌트 기반 개발(CBD), 소프트웨어 아키텍처(Software Architecture), 데이터 마이닝(Data Mining), 지식공학 (Knowledge Engineering)

김 수 동(Soo Dong Kim)

[정회원]



- 1981년 3월 ~ 1984년 2월 : Northeast Missouri State University 전산학 (학사)
- 1988년 2월 ~ 1991년 2월 : The University of Iowa 전산학 (석사/박사)
- 1991년 3월 ~ 1993년 2월 : 한국 통신 연구개발단 선임연구원
- 1994년 3월 ~ 1995년 2월 : 현대전자 소프트웨어연구소 책임연구원
- 1995년 3월 ~ 현재 : 숭실대학교 컴퓨터학부 교수

<관심분야>

서비스 지향 아키텍처(SOA), 클라우드 컴퓨팅(Cloud Computing), 모바일 서비스(Mobile Service), 객체지향 S/W공학, 컴포넌트 기반 개발(CBD), 소프트웨어 아키텍처 (Software Architecture)