Knowledge Validation Using Objectivity and Corroborativeness of Web Resources

Seongchan Kim, Jiyeon Choi, and Mun Y. Yi

Dept. of Knowledge Service Engineering, KAIST, South Korea
`sckim, jeeyeon51, munyi@kaist.ac.kr`

Abstract. In this study, we propose a method to validate knowledge candidates using the Web to minimize the false positive rate in a knowledge base (KB). Our approach assesses the objectivity and corroborativeness of a triple, which is the basic form of knowledge, using diverse Web resources. Compared to the state-of-the-art baseline of the Defacto framework, our approach demonstrates superior false positive rates, enabling more effective filtering of false triples in the construction of a KB.

Keywords: knowledge validation, Web-based validation, objectivity, corroborativeness

1 Introduction

Knowledge validation that identifies the truth value (true or false) of facts is a vital step for constructing a reliable, usable knowledge base (KB). To the greatest extent possible, true facts should be distinguished and stored in a KB while incorrect or unreliable facts are filtered out as the existence of false triples in a KB can cause serious problems for those applications that use the KB. For instance, a wrong answer can be generated from a Q & A system that references a KB implanted with false knowledge. Therefore, techniques for validating knowledge are essential for the use of KB-based systems.

In literature, several studies have attempted to validate facts. *Defacto* demonstrated a new approach that allowed testing whether a given fact (i.e., an RDF triple) [1] could be trusted. More specifically, *Defacto* takes a statement from an RDF as input to the Web and tries to find evidence for validation using webpages. It combines the trustworthiness of a Web resource and textual evidence for validating triples with a machine learning technique. On the other hand, there has been a study about a system - *Honto? Search* for helping users determine the trustworthiness of uncertain facts considering their sentimental aspects [2]. The semi-automatic system incorporates what sentiments are mentioned on the relevant webpages.

In this paper, we present a new technique to utilize the objectivity and corroborativeness of Web resources for RDF truth validation. For validating an RDF triple, we identify confirming sources on the Web as introduced in *Defacto* [1]. On top of that, we estimate to what extent the confirming sources is neutral

by performing the sentiment analysis of the sentences in the page. Contrary to *Honto? Search* [2], our approach automatically utilizes the results of sentiment analysis to truth validation. Furthermore, we check the corroborativeness of the RDF - the extent to which different types of evidence support the truthfulness of the triple. We count various types of Web resources such as images, videos, and news for validation. Finally, we evaluate the performance of the proposed approach by comparing it with *Defacto*, which serves as a baseline.

2   Approach

Our primary purpose is to determine whether an RDF triple (s, p, o) given is true or false. We deal with this problem as a binary classification. In this section, we describe how we estimate the objectivity and corroborativeness of webpages. Figure 1 shows a complete picture of our strategy for estimating the proposed features of webpages retrieved for a given triple t (s, p, o). The estimated objectivity features are replaced with the trustworthiness features proposed by *Defacto* [1] and the corroborativeness features are added to the *Defacto* framework[1].
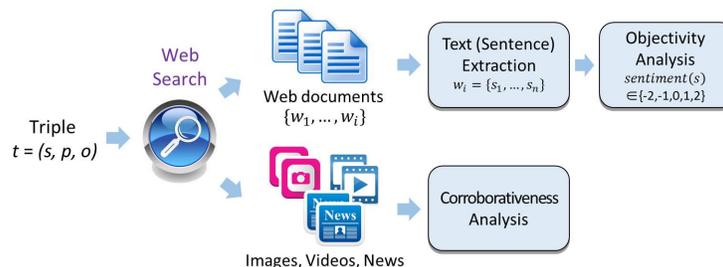

Fig. 1: Overview of objectivity and corroborativeness analysis for Web resources

The objectivity of a webpage is calculated by measuring the degree of sentiments of all sentences in the document, and it is estimated as follows:

$$obj_{sum}(w) = \sum_{s \in w} \frac{2 - |sentiment(s)|}{2} \qquad obj(w) = \frac{obj_{sum}(w)}{n}$$

where t is a triple, w is a web document retrieved by t, w consists of a set of sentences s and is denoted as $w = \{s_1, s_2, ..., s_n\}$, where n is the number of sentences. The sentiment value of a sentence takes one of the following values: $sentiment(s) \in \{-2, -1, 0, 1, 2\}$ (very negative, negative, neutral, positive, very positive). Therefore, $obj(w)$ has a range from 0 (subjective) and 1 (objective). After obtaining the objectivity score, we multiply it with trustworthiness score and textual proof score of w considering that webpages with high trustworthiness, proof score, and objectivity increase the confidence in the input fact.

---

[1] http://aksw.org/Projects/DeFacto.html

$$F_{fobjsum}(t) = \sum_{w \in s(t)} (f(w) \cdot scw(w) \cdot obj(w)) \quad F_{fobjmax}(t) = \max_{w \in s(t)} (f(w) \cdot scw(w) \cdot obj(w))$$

$f(w)$ is instantiated by three trustworthiness scores: topic majority (tm_{web}), topic majority in search results (tm_{search}), and topic coverage (tc) of w. $scw(w)$ is the proof score on a webpage w. The proof, one of the supporting evidence on webpage that the triple given is true, is defined as a textual occurrence between s and o within a certain token distance. The details for the trustworthiness and proof score are described in [1]. We generate the objectivity features multiplying three criteria using their sum and maximum as different features. Finally, we propose six objectivity features by a combination of three trustworthiness measures (tm_{web}, tm_{search}, and tc) and two cases (sum and max): `TM_Web_Obj_Sum`, `TM_Web_Obj_Max`, `TM_Search_Obj_Sum`, `TM_Search_Obj_Max`, `TC_Obj_Sum`, and `TC_Obj_Max`.

Moreover, for validation of RDF triples, we utilize various web resources: images, videos, and news. We counted the number of hits of images, videos, and news. Our rationale is that true triples are more likely to have more supporting evidence in multiple forms (i.e., images, videos, and news) than the false ones because diverse types of information about the triples is accumulated on the Web. For example, if we have an RDF triple such as (Maroon 5, Song, Sugar) and search the Web with the "AND" operator (e.g., "Maroon 5" AND "Song" AND "Sugar") using the Bing Search API[2], the API returns 14200 images, 833000 videos, and 16800 news as of June 2015; however, with a false triple ("Maroon 5", "Song", "Blank Space"), the API returns 4530, 103000, and 4290, respectively. ("Blank Space" is a song by Taylor Swift.) For corroborativeness, we employ the following three features: `Num_Images`, `Num_Videos`, and `Num_News`.

3 Experiment

We collected 570 RDFs from DBpedia as true triples on the top 57 most frequently used properties in DBpedia[3]. We randomly selected triples containing the property. We derived the false triples from the true triples with the following restriction. A triple (s', p', o') is generated, where s' and o' are randomly selected resources, and p' is a randomly selected property from the defined properties. We applied the *Defacto* as a baseline, which is the only Web-based RDF true/false validation framework, to our best knowledge. To measure the objectivity of sentences in webpages, we used a sentiment analysis tool[4] in the Stanford CoreNLP. The classification of the two classes (true and false) was conducted using ten-fold cross-validation. We used random forest (RF), which have been widely adopted for classification, in the Weka toolkit[5] with the default parameter values given in Weka. We report on three measures: precision (P), recall (R), F-1 score (F1) (micro-averaged), and false positive rate (FP rate).

The classification results are shown in Table 1. Note that the performance was measured by replacing the six objectivity features with those proposed by *Defacto* (e.g, from `TM_Web_Sum` to `TM_Web_Obj_Sum`) and adding the corroborativeness features. A 5.9% and 6.7% decrease in the FP rate with RF was achieved using the objectivity and corroborativeness features, respectively.

Table 1: Performance of classification by adding group of features over baseline

	Random Forest			
	P	R	F1	FP rate
Defacto (Baseline)	0.761	0.761	0.76	0.239
+Objectivity (Replacing)	0.776	0.775	0.775	0.225
+Corroborativeness	0.791	0.79	0.79	0.21

Mean objectivity score of the retrieved websites from the true triples was 0.871 (StdDev: 0.202) and score from the false was 0.804 (StdDev: 0.262). This indicates that the websites from the true triples have less sentiments on their texts than those of the false. The overall performance decrease in the FP rate and the different distribution of objectivity score in the two classes confirm the effectiveness of objectivity in judging the trustworthiness of the fact. In addition, our results are in general agreement with the results of a prior study, in which users admitted that sentiment analysis for a fact is useful in fact validation [2]. Furthermore, corroborativeness is shown to be highly effective, supporting the rationale that diverse types of information, not limited to the text, accumulated on the Web are useful evidence about the truthfulness of the given triple.

4 Conclusion and Future Work

In this paper, we presented an approach for applying the objectivity and corroborativeness of webpages retrieved by the RDF triples in the true/false validation of given RDF triples, showing the effectiveness of these concepts in knowledge validation. For future work, we are planning to extend the current study to perform a deeper analysis about the differences of sentiment distributions for true and false fact validation as well as the confirmation of our approach even in political triples that could have limitations.

5 Acknowledgement

References

1. Lehmann, J., Gerber, D., Morsey, M., Ngonga Ngomo, A.: DeFacto - Deep Fact Validation. In: ISWC2012, 312–327 (2012)
2. Yamamoto, Y., Tezuka, T., Jatowt, A., Tanaka, K.: Supporting Judgment of Fact Trustworthiness Considering Temporal and Sentimental Aspects. In: WISE2008, 206–220 (2008)