Handling Uncertain and Ambiguous Spatial Expressions in Text Using Fuzzy Logic

V. R. KANAGAVALLI

Research Scholar, Sathyabama University
Chennai, Tamil Nadu, India
Kanagavalli.teacher@gmail.com

Dr. K. RAJA

Dean (Academics), Alpha College of Engineering
Chennai, Tamil Nadu, India
raja_koth@yahoo.co.in

Abstract

The knowledge era, as this era is called, poses the challenge of churning knowledge from the pool of information available from various sources. Much of the text documents acting as an information source and the query posed by the user implicitly have a geographic or spatial reference component present in it. This logically leads to the conclusion by the previous studies that more than 80% of the searches are pertaining to geographic locations. Text documents imply the usage of natural language and as such it yields to explicit vague fuzzy descriptions involving linguistic terms such as near to, far from, to the east of, very close and also implicit vague spatial references. Any text document which contains physical location specifications such as place names, geographic coordinates, landmarks, country names etc., are supposed to contain the spatial information. Fuzzy logic is an extension to the Boolean crisp logic to accommodate for the fuzziness of an element belonging to a set. The understanding and extraction of spatial components is a primal area of study not only in the field of information retrieval but also in various other fields such as Robotics, Psychology, Geosciences, Geography, Political Sciences, Geographic Economy, Environmental, Mining and Petroleum Engineering, Natural Resources, Epidemiology, Demography etc., Given a query involving events, the aim of this ongoing research work is to extract both the explicit and implicit spatial references from the text documents using fuzzy logic techniques.

***Keywords*: Spatial uncertainty, Fuzzy logic, Spatial cognition.**

1. Introduction

The sources of information in the knowledge era are images, web pages, audio files, video files, spread sheets, maps and text documents that exist both in structured and unstructured format. The information can be categorized in to explanation of concepts, description of products and narration of events which may be disseminated impromptu by the person or as a reply for a query in a question answer based systems like blogs or discussion forums. Also there is plenty of information in social sites in various formats as mentioned above. In spite of the multiple types of information, text dominates the field since much of the information is present either in structured or unstructured text form. There is a repository of information in text corpus in the form of journals, encyclopedia articles, books, technical reports, environmental impact reports, laws and legislation. Much of the information consists of spatial components as part of it. For instance the happening of an event is always tied up to a spatial location. Also the description of concepts or products or natural entities may contain spatial reference to the same. Also the resource finding from tourist information reports has spatial component present in the. Not only the description of events, but various scientific documents and historical documents are also replete with spatial descriptions in them Understanding the spatial references in text would be applicable in majority of areas like disaster management, finding resources in tourist reports, analyzing the geographical scope of the document in domains like environmental preservation, laws and legislation. The spatial terms in a document may consist of statements like in front of, at the back side of, to the left of, to the right of, close to the east of Estancia, quite near the bus stop, far away from the residential area, in and around Chennai, behind the Nehru Park, Just before, Few steps away from, at the stone's throw away etc.,

The spatial information can be orientation information, distance information, neighborhood information and topological information. Geographical information retrieval has been an active area of research for a long time now and has been applied in multiple disciplines like GIS, transportation analysis, archaeological events etc.,[1][2][6].

2. Uncertainty

Uncertainty arises whenever we lack the certainty of existence or occurrence or happening of an object or an event. Uncertainty propagates either in collection of data or expression of data. In this work we are concentrating in the uncertainty arising during the expression of explanation of object or an event. There are two types of uncertainty, namely objective uncertainty and subjective uncertainty. The objective uncertainty cannot be eliminated completely even after adding more information to it. The subjective uncertainty can be reduced by adding more information to it since it arises due to the lack of knowledge whereas the former arises due to the presence of inherent randomness.

There is an inherent uncertainty in the spatial references found in the text document. The uncertainty may arise due to the user's perception of the distance, the qualitative spatial information provided in the natural language description instead of quantitative spatial information. Also the spatial information found in natural language or free form text reports are ambiguous with homonyms embedded in it. The identification of the spatial component in the text relies upon the available information or the knowledge base, especially in the case of special names given to certain spatial locations (City of temples, Capital of Tamilnadu, uranga nagaram (Madurai), monument of love in the city =' tajmahal' if the city of interest or discussion is Delhi) which encompasses the user's knowledge about the locality and also highly context dependent. These implicitly mentioned spatial locations poses challenge to the geo indexing process.

2.1. Methods handling uncertainty

There is inherent uncertainty in natural language used for providing spatial information such as route information for robots, tourists, location based services like mobile applications etc. There are various research works which deal with this issue to find geospatial reasoning and to understand natural language directions [7] [8] [9]. The spatial information is converted into geographic coordinates for resolving ambiguity since geographic coordinates to a place is unique and bypasses the homonymy, synonymy inherent in the spatial expressions found in text document [10]. Berztiss outlined common methods of uncertainty management, including Bayesian inference, fuzzy sets, fuzzy logic, possibility theory, time petri nets, evidence theory, and rough sets [18]. Walley compared four measures of uncertainty in expert systems including Bayesian probabilities, coherent lower and upper precisions, belief functions of evidence theory, and possibility measures in fuzzy theory [19]. Klir studied uncertainty and information as a foundation of generalized information theory [20]. A summary of the most used techniques is presented by the authors of [17]. The theories most used are probability theory, fuzzy theory, derived uncertainty theory and info-gap theory. Probability theory handles objective uncertainty and subjective uncertainty using statistics. Monte Carlo method, Bayesian method, and Dempster- Shafer evidence theory are the methods extended from the existing probability theory. Fuzzy theory handles subjective uncertainty at ease. Ambiguity arising out of natural language expressions can be handled through fuzzy sets with membership functions.

3. Methods dealing with handling spatial term extraction

To retrieve spatial expressions from text documents the basic activities performed are indexing, querying, comparison, and feedback.

The major text processing steps include stemming, lemmatization, part of speech tagging (POS), syntactic analysis and Named Entity Recognition (NER). This author's previous work [15] proposes a method for detecting the presence of spatial uncertainty in the text and dealing with spatial ambiguity using named entity extraction techniques coupled with self learning fuzzy logic techniques. The other techniques used are Machine learning techniques, Geographic thesaurus and gazetteer, Texts geo-tagging and geo – footprints and Geographic ontology. The issue with spatial cognition through information retrieval from text documents is that it is highly domain specific since the gazetteer and thesaurus that can be designed cannot be all encompassing and exhaustive in nature [5]. The authors of this work propose the use of fuzzy logic for effective retrieval of ambiguous spatial terms, spatial expressions from the text and also modeling the same.

4. Fuzzy Logic

Fuzzy theory, proposed by Zadeh in 1965 [16], is another good way to deal with vagueness uncertainty arising from human linguistic labels. It provides a framework for modeling the interface between human conceptual categories and data, thus reducing cognitive dissonance in problem modeling so that the way that humans think about the decision process is much closer to the way it is represented in the machine. The concept of fuzzy set extends the notion of a regular crisp set and expresses classes with ill-defined boundaries such as young, good, and important, etc. Within this framework, there is a gradual rather than sharp transition between non-membership and full membership. A degree of membership in the interval [0, 1] is associated with every element in the universal set X. Such a membership assigning function ($\mu A: X \rightarrow [0, 1]$) is called a membership function and the set (A) defined by it is called a fuzzy set [4]. There are works that use fuzzy logic for handling uncertainty in GIS [11]. These works concentrate on handling the spatial uncertainty that creeps in to the system

while data acquisition and data representation. Also, fuzzy logic is adept at handling the imprecise boundary that is inherent in spatial objects extracted from the satellite images and maps.

The ontology based methods cannot resolve deictic references which involve the observer's point of view in the narration. Also the traditional Boolean logic system is inefficient and insufficient in handling the uncertainty and ambiguity associated with natural language descriptions of events involving spatial descriptions that is based on the human cognition and perception of the event. The qualitative representation can be converted to quantitative representation using fuzzy membership functions. Accessing information at multiple granularity levels can be used in solving user requirements at specific levels. There are works that deal with granularity of objects, locations and actions that are embedded in an event [12][13][14]. The granularity at varying levels can be achieved by using fuzzy sets. The degree of membership of a location to a known or indexed location can be easily determined using fuzzy membership values assigned to the spatial terms extracted from the text. Fuzzy sets also enable to understand the spatial context at multiple granularity levels, by understanding the type of specified place, specific building which is achieved by framing more rules or membership functions.

5. Proposed Work

The figure given below (Figure 1) shows the proposed architecture depicting the sequence of steps taken to resolve the uncertainty and ambiguity in the text documents. The text corpus is treated as an input to the system from wherein the spatial references are extracted. The spatial references are partially disambiguated using the manual annotation using the gazetteer and then the most possible spatial references are fed into the next module which interacts with the fuzzy rule base wherein the granularity is fine tuned and the possible locations are extracted from the document set and displayed to the user through an interface.

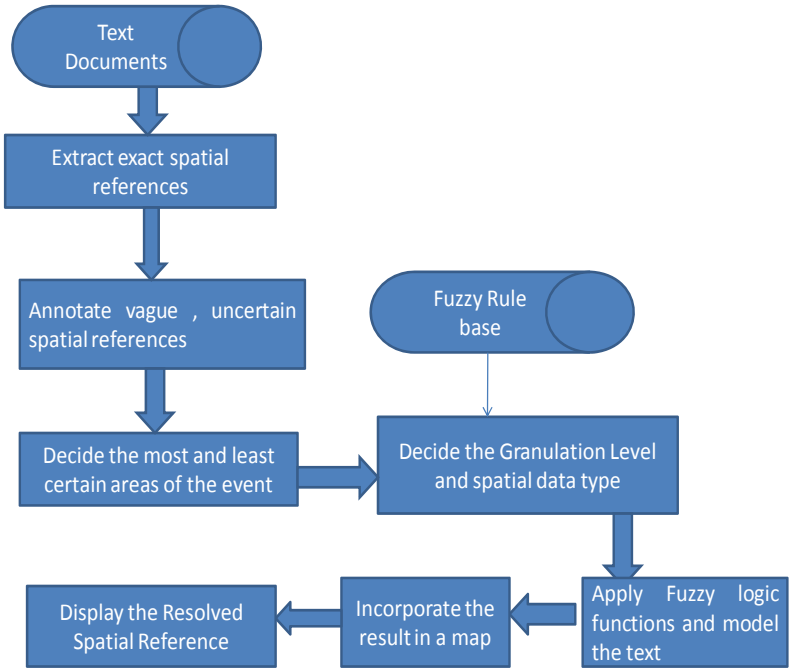

Fig 1: The proposed architecture of the system

The spatial information searched for generally includes the spatial relationships of geometrically defined spatial entities. The spatial information retrieval from text would be find out which of the documents in the text corpus has references about a spatial point/ region of interest. Resolving the uncertainty inherent in spatial expressions and spatial references in text documents can be used to provide an efficient and effective question answering systems, discourse analysis, disaster management applications, mobile robotics, finding resources in tourist applications, GIS etc.,

The extraction of spatial references is fine-tuned using fuzzy rules. There is a gazetteer involving locations of a limited area. Identifying the spatial references from the text document is achieved by comparing each term with the built- in gazetteer and by referring the spatial adjectives with each term. The authors of this paper assign a fuzzy membership values to the spatial terms indicating their belonging to the gazetteer. If there are adjectives that belong to the spatial property of the term, such as near to, far away from, to the north of, then the term is selected as a spatial term. If it is not present in the thesaurus currently then it is added to the thesaurus indicating

the presence of one more spatial term, thus the thesaurus keeps growing with spatial terms encountered with each document. The user is given a choice to confirm whether this is a valid spatial term or not. In case of any modification, it can be done manually also for individual entry. Though this is feasible in the early phases, as the thesaurus is expected to grow much in size, this is not advocated for.

Once the geographic terms are retrieved they have to be assigned fuzzy membership values which determine the degree of possibility that the spatial location belongs to the selected geographic coordinate system. The ambiguity in place names is resolved by assigning membership values to the different coordinate systems to which the place may belong. The fuzzy membership value is assigned by referring to the other spatial terms associated with the document.

6. Conclusion and Future work

The authors are retrieving the documents from the text corpus which are relevant to the spatial queries. The order and ranking of the documents retrieved are different from the traditional Boolean ranking since the documents are ranked on the basis of relevance using fuzzy logic techniques. The fuzzy membership functions determine the spatial relevance of the documents and the fuzzy rules decide the relevance. The spatial similarity between two documents is also evaluated on basis of the fuzzy rule base. The granularity of the query and the spatial information present in the text are used to resolve the uncertainty of the spatial information. Granularity can be adjusted as per user requirements by modifying or adding more rules using the user interface provided in the proposed system. Possibility functions, Fuzzy logic techniques are used to model the uncertainty of the spatial information present in the text instead of the probability logic.

The limitations of this work is that it would answer spatial queries involving spatial attributes only and not spatial queries involving geometric shapes since it is querying the textual data and not the spatial database. Also, the proposed system handles only point in polygon queries, region are answered. The future work of the author involves solving the multimedia queries to an extent. The path queries cannot be handled since it requires network information which is generally not found in text corpus. Distance and buffer zone queries also require information that is usually not found in the text corpus.

References

[1] C. B. Jones, Ross. S. Purves, 2008. Geographical Information Retrieval, International Journal of Geographical Information Science, 22(3)

[2] S. Kikuchi et al., Place of possibility theory in transportation analysis. Transportation Research Part B 2006. Elsevier

[3] Rock, Nathaniel Robert."Mapping geospatial events based on extracted spatial information from web documents." master's thesis, University of Iowa, 2011.http://ir.uiowa.edu/etd/1068

[4] George J. Klir and Bo Yuan. Fuzzy sets and Fuzzy logic, Theory and applications,

[5] Debra, Rajiv Chopra, Rohini Srihari. Domain Specific Understanding of Spatial Expressions. citeseerx.ist.psu.edu / viewdoc / download ?doi=10

[6] Kate Byrne and Ewan Klein.Automatic Extraction of Archaeological Eventsfrom Text. May 2009.

[7] Hans w. guesgen. Reasoning About Distance Based on Fuzzy Sets. Applied Intelligence 17, 265–270, 2002

[8] Thomas Kollar et al., Toward Understanding Natural Language Directions. Naval Research

[9] Geospatial reasoning in a Natural Language Processing (NLP) Environment. Bitters B.

[10] Kalev H. Leetaru Fulltext Geocoding Versus Spatial Metadata for Large Text Archives: Towards a Geographically Enriched Wikipedia . D-Lib Magazine September/October 2012 Volume 18, Number 9/10 doi:10.1045/september2012-leetaru.

[11] H.W. Guesgen, J. Albrecht. Imprecise reasoning in geographic information systems Fuzzy Sets and Systems 113 (2000)

[12] Mulkar-Mehta, R.; Hobbs, J. R.; and Hovy, E. 2011. Granularity in Natural Language Discourse. International Conference on Computational Semantics, Oxford, UK 360—-364

[13] Mulkar-Mehta, R.; Hobbs, J. R.; and Hovy, E. Applications and Discovery of Granularity Structures in Natural Language Discourse. Logical Formalizations of Commonsense Reasoning — Papers from the AAAI 2011 Spring Symposium (SS-11-06)

[14] V. R. Kanagavalli, K. Raja, Graduated granulation of spatial information for efficient, effective business activity monitoring, Fuzzy Sets and Systems, pp 99-101. 2010

[15] V. R. Kanagavalli, Dr. K. Raja, Detecting and resolving spatial ambiguity in text using named entity extraction and self learning fuzzy logic techniques, NCT2DS, 2011.

[16] L. A. Zadeh, “Fuzzy sets,” J. of Information and Control, vol. 8, pp. 338–353, 1965.

[17] Yiping Li, Jianwen Chen, and Ling Feng Dealing with Uncertainty: A Survey of Theories and Practices, IEEE, 2012

[18] A. Berztiss, Uncertainty Management, University of Pittsburgh, Department of Computer Science, Pittsburgh PA 1526 USA, 2002.

[19] P. Walley, “Measures of uncertainty in expert systems,” Artificial Intelligence, vol. 83, pp. 1–58, 1996.

[20] G. Klir, Uncertainty and Information, Foundations of Generalized Information Theory. Hoboken: Wiley, 2006.