Ontology Synthesis & Generation Using AI Orientated Hybrid Learning for Microeconomics

[1]Abhijith Roy, [2]Gerard Deepak and [3]Santhanavijayan A
[1,2,3]Department of Computer Science and Engineering
[1,3]National Institute of Technology, Tiruchirapalli, India
[2]Manipal Institute of Technology, Bengaluru, Manipal Institute of Higher Education, Manipal, India
[2]gerard.deepak.christuni@gmail.com

Abstract. There is a dearth for a strategic automatic framework for ontology synthesis and generation which encompasses semantics oriented reasoning for a domain of importance such as microeconomics. This research paper proposes ontology synthesis and generation model that encompasses hybrid learning paradigms for microeconomics as a prospective domain. The dataset of documents relevant to microeconomics as a domain is subjected to classification using the LSTM classifier and models such as TF-IDF, Structural Topic Modeling are encompassed to generate the initial seed knowledge from the dataset of documents. Standard strategic knowledge store repositories such as YAGO and RDF generation helps in adding to the auxiliary knowledge of the model. The exponential increase of the knowledge is achieved through metadata generation and making it more atomic into the model through the classification by the convolutional neural networks (CNN). Semantics oriented reasoning is achieved through standard strategic semantic similarity measures such as NPMI, Associate PMI, Second-Order co-occurrence PMI, Jaccard Similarity and intermediate optimization of the results is achieved using particle swarm optimisation (PSO) to be the best in class strategic entities as an ontology. The proposed framework is the first in class framework that formalizes a ontology for microeconomics as a domain and achieves a precision of 95.09% with a F-measure of 96.957% and an FDR value of 0.05, thus proving to be the best in class model for ontology generation.

***Keywords*:** Ontology Synthesis, TF-IDF, JACCARD Similarity, RDF, Second-Order PMI, LSTM, YAGO, CNN

1. Introduction

An ontology serves as a hierarchical arrangement of lexical terms, along with their syntactic and semantic connections, aimed at constructing a structured knowledge framework. This challenge is being addressed through the utilization of natural language processing methods and the creation of ontologies from natural language text. These ontologies, once established, contribute significantly to enhancing information retrieval capabilities. Extensive research has been conducted within the realm of text processing and the automated construction of ontologies from textual data to confront these complexities. Microeconomics has been long neglected in the field of formal ontology and as it has gained its momentum in the past decade, there is an urgent need to develop a common language between domain experts, economists, and ontology engineers. Governments and Education rely on microeconomic strategies to build policies and laws to build effective training and policies that shape a country. This paper delves into the realm of Ontology Synthesis and Generation for the domain of Microeconomics, employing AI-oriented hybrid learning techniques to navigate these intricate processes.

Motivation: The ultimate motivation of the proposed strategy is looking for a strategic model for ontology synthesis which is automatic in nature and rare domains such as microeconomics also require formal ontology as they are gaining importance and momentum in the present day and age. Most importantly, currently the World Wide Web is transforming into Web 3.0, the ultimate need for a strategic ontology is required as it can accelerate and organize the content of the Web 3.0 deep in its structure. Owing to this, there is an automatic model which encompasses a semantic intelligence for ontology generation for the domain of microeconomics has been proposed in this paper.

Contribution: The contributions of the proposed framework include the usage of TF-IDF and Structural Topic Modeling (STM) which ensure to initiate the seed entities from the dataset of documents comprising microeconomics domain and enrichment through standard APIs such as WikiData, YAGO knowledge stores, and generation of triadic knowledge through RDF which in turn generates the metadata which is further classified using convolutional neural networks (CNN). Subsequently, metadata is generated through the entities that are classified through the LSTM classifier and furthermore, the entities which are common and convergent come out of the RDF and the LSTM is also used to produce metadata is further classified using the CNN classifier. So convergence of entities and semantic oriented reasoning is achieved through semantic related similarity measures such as Associate PMI, NMPI, Second-Order PMI, and Jaccard Similarity. Particle Swarm Optimisation model is the intermediate metadata optimisation to refine the initial attainable solution into a much more optimal solution set to yield the best in class entities as ontology synthesis further formalize axiomatization and reason in order to make it a formal ontology. Average precision and average recall percentage of the proposed model is improved and the FDR measure is decreased when compared to all the baseline models.

2. Related Works

Tijerino et al. [1] provides an approach towards ontology generation from tabular data while exploring the conversion of structured data into ontological structures, which proves to be a valuable method for integrating data from diverse sources. Wang et. al [2] focuses on formal ontology generation using deep machine learning, whilst primarily highlighting the application of advanced machine learning techniques for constructing formal ontologies. Kumar et al. [3] present an automated method for ontology curation from plain text using proven statistical and NLP methods, emphasizing the extraction of concepts from unstructured textual data. Bergamaschi et al. [4] provides an insight on the semantic integration of both structured and semistructured data sources, offering a unique perspective on how different data formats can contribute to ontology synthesis. Biskup et al. [5] propose an approach to extract information from varied sources using ontologically specified target views, providing strategies for harmonizing information from diverse domains. Castano et al. [6] discusses conceptual schema analysis techniques contributing to understanding how to identify and formalize concepts for ontology. Chiang et al. [7] discusses reverse engineering of relational databases, offering methodologies for deriving an EER model from an existing database, which could aid in ontology construction. Poelmans et al. [8] explore practical implications of applying formal concept analysis to knowledge processing, showcasing potential scenarios where this approach can be utilized effectively in ontology synthesis. Godin et. al [9] introduces algorithms for incremental concept formation based on Galois lattices, presenting strategies for gradually building ontologies. Cocchiarella [10] introduces conceptual realism as a structured ontology, enhancing our comprehension of the fundamental principles and underpinnings of ontology development.

3. Proposed System Architecture

The proposed system architecture of ontology synthesis and generation framework utilized AI-focused Hybrid learning within the selected domain of microeconomics as shown in Fig.1. This is the first of its kind ontology generation using automation principles for microeconomics as a prospective domain, here the dataset of documents is subjected to term frequency-inverse document frequency (TF-IDF) model where the informative terms on the documents based on the concept of rarity of frequency of terms within the documents corpus is yielded. Once the TF-IDF discovered terms are obtained from the dataset, these terms although informative in nature, they can be contextualized for which topic discovery and modeling is required which is achieved using structured topic modeling. Structured topic modeling is the topic modeling strategy of a framework that assumes the web as the reference corpora and topics which are hidden are discovered and these newly discovered topics are used to contextualize the terms further. These topics discovered through STM are fed into the WikiData API to yield community contributed verified cloudsource open linked data which is hyperlinked in nature and it is accommodated on a WikiData pipeline which is further fit into YAGO knowledge base, which is a knowledge store repository through SPARQL query via an agent designed using JADE.The entities that come out of the YAGO pipeline are extensively large and these entities are aggregated with that of those discovered with the help of STM, TF-IDF and so a contextual knowledge tree is formulated. For each of the instances of the context knowledge tree, RDF is generated.RDF is a Resource Description Framework which loads triadic entities in their triple X structure which is a form of Subject-Predicate-Object (S-P-O). However, considering that the Predicate acts as linkage connecting the Subject and object, due to heterogeneity of predicates and to also avoid redundancy, the predicates are dropped and subject and object of the RDF is retained alone and RDF is generated using a Photo RDF-Gen.In this process, the Photo RDF-Gen where the predicates are dropped and the subject and object are alone retained. For each of the RDF generated entities, subject and object together form the metadata which is generated using a tool named DSpace.

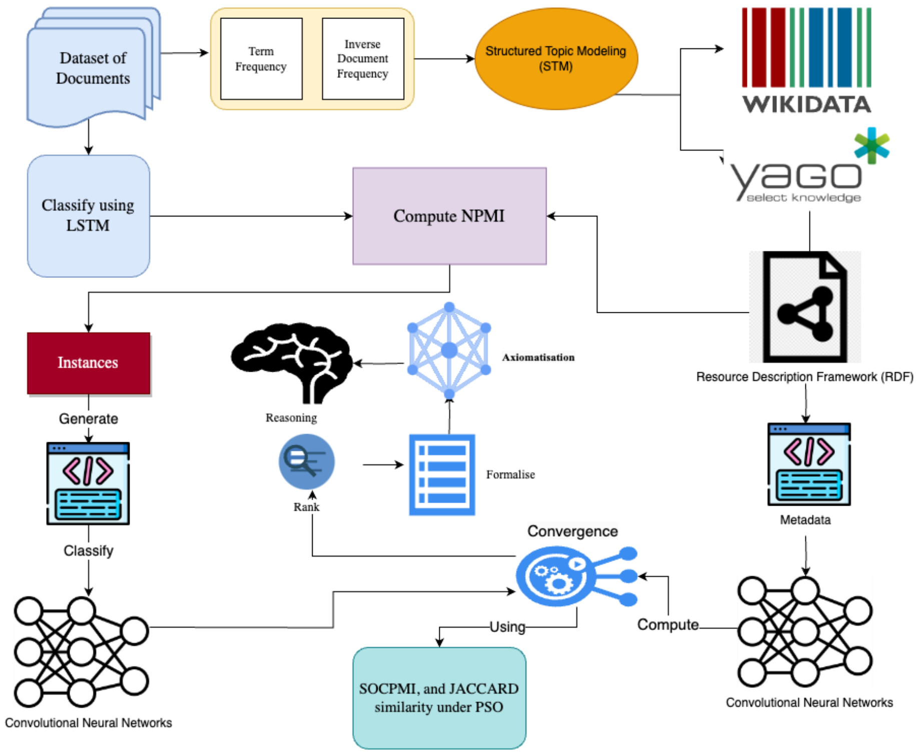

Fig. 1. Proposed System Architecture for OSGAM

The metadata generated using DSpace is exponentially large, therefore, it needs to be handled in the localized framework and has to be made permeable so therefore classification is accomplished employing the potent CNN classifier, a deep learning classifier, owing to its strength and since it is an implicit selection model, the CNN is made use of and the entities which come out of the CNN pipeline are further use of the model. Subsequently, the LSTM, another robust deep learning classifier, is employed to categorize documents from datasets. The LSTM is used because it also works on implicit feature dataset selection and works well over a document dataset. Relevance computation through NPMI is achieved between the generated RDF and the entities that come out of the LSTM. NPMI is computed with a threshold of 0.60 to increase the relevance and positive values of NPMI between 0-1 alone are considered to be the intermediate instances. The intermediate instances are highly relevant but shallow so they are further enhanced by generating the metadata.

At this juncture, a tool named RepoMMan is harnessed to augment the count of instances as well as the diversity of the metadata so henceforth a different tool is used. The entities that come out of the metadata are classified using the CNN in order to make it more atomic, permeable to the model and handleable due to the large scale of the metadata. So entities which come out of the CNN classifier from the metadata through the NPMI and the entities which come out of the CNN through the RDF fed generated metadata are from the two distinct metadata tools, RepoMMan and Dspace. Both the outcome of the classified instances are subjected to computation of conversions achieved using SOCPMI, associate PMI and JACCARD similarity. The SOCPMI is said to be step deviance of 0.60 owing to the strength of associate PMI and JACCARD is said to have a step deviance associate 0.75 because associate PMI itself is very stringent to compute the initial solution set. The initial set of solutions already encompasses a substantial number of entities, subsequently, optimization is pursed until all entities are exhausted using the particle swarm optimization algorithm. This algorithm is a meta optimisation approach utilizing the JACCARD similarity with the same threshold as the objective function, yielding entities. The resulting entities are then ranked in ascending order of JACCARD similarity, formulated on the basis of JACCARD similarity, creating an edge between highly similar terms and the former ontology is modeled. This ontology which is modeled is subjected to Axiomatization using an agent where 'is-a-part-of', 'has-a-part-of', is a subclass of generalization, composition, and specialization relationships are induced. The axiomatized entities are a formal ontology which is reasoned out using the Pelitte and Herman reasoner so this ensures a best fit ontology for which instances can be done through Google's knowledge graph API and fed it as a single large knowledge graph with instances. The second-order co-occurrence Pointwise Mutual Information (PMI) is employed to gauge the statistical relationship between two item pairs within a text data corpus. It helps to determine how often two pairs of items appear together compared to what would be expected by chance. The formula governing the calculation of second order co-occurrence PMI is depicted in Equation (1).

$$f^{pmi}(t_i\,, w) \;= log_2 \frac{f^b(t_i\,,w) \times m}{f^t(t_i) \times f^t(w)} \quad (1)$$

The first term f^{pmi} signifies the PMI score between a focal term and its contextual term, taking into account frequency weighting. The second term, $f^b(t_i\,, w)$ refers to the frequency of the joint occurrence of the focal term and contextual term within the corpus, reflecting the number of times they appear in conjunction. The first term on the denominator, $f^t(t_i)$ indicates the frequency of the target term in the corpus. It indicates how many times the target term appears overall.The second term on the denominator, $f^t(w)$ indicates the frequency of the context term in the corpus. It indicates how many times the context term appears overall. m is the smoothing factor or constant that is added to the numerator of the equation. It helps undefined values when there are no co-occurrences of the target term and context term in the corpus. Jaccard similarity coefficient as depicted in Equation (2), which is a measure of the similarity between two sets, A and B. This statistical measure is utilized to assess the comparability and variety of sets within a sample.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \qquad (2)$$

J(A,B) represents the Jaccard similarity coefficient between sets A and B. This coefficient quantifies the degree of overlap or similarity shared by the two sets, yielding a value ranging from 0 to 1. A value of 0 pimples no overlap =, whilst a value of 1 signifies complete overlap or similarity. |A∩B| denotes the intersection of sets A and B, indicating the count of elements that exist in both sets A and B.

|X∪Y| indicates the cardinality or size of the union sets X and Y. It represents the overall count of unique elements encompassed by both sets A and B.The Term Frequency (TF) is depicted in Equation (3) serves to quantify the frequency of a particular term within a given document, thereby assessing its relevance and significance within that context.

$$tf(f,d) = \frac{f_{t,d}}{\Sigma_{t\prime \epsilon d} f_{t\prime,d}} \qquad (3)$$

By dividing the occurrences of the term within the document by the total term count, the TF equation standardizes the term’s frequency in relation to the document’s content. This normalization accounts for differences in document lengths, ensuring that longer documents don't have an inherent advantage due to having more terms. Conversely, the Inverse Document Frequency (IDF) equation gauges the term’s significance or rarity across a collection of documents. It quantifies the discriminative power of a term by measuring how much information it provides in distinguishing a particular document from the entire corpus. The formula for inverse document frequency is as follows:

$$idf(t,D) = log \frac{N}{|\{d \in D: t \in d\}|} \qquad (4)$$

Corpus Size: This denotes the overall count of documents within the corpus, signifying the entirety of the text collection. Term Occurrence Frequency: This quantifies how many documents within the corpus include the specific term. The IDF is depicted in Equation (4) adjusts the significance of terms based on their occurrence across documents by taking the logarithmic ratio between the total number of documents in the corpus and the number of documents containing the term. This approach mitigates the influence of frequently appearing terms and enhances the significance of rarer terms in the context of the corpus.CNN is an acronym for Convolutional Neural Network, a sophisticated architecture within the realm of deep learning. This architecture is extensively employed in diverse computer vision tasks, encompassing image classification, object detection, and image segmentation. CNNs excel at autonomously discerning and extracting intricate features from input data, particularly images, by harnessing the potency of convolutional layers.

It's important to note that while CNNs can provide valuable contributions to ontology engineering, they are typically part of a broader set of techniques and methodologies employed in this domain. The utilization of CNNs can be combined with other ontology engineering practices, such as ontology modeling, reasoning, and validation, to ensure the overall quality and effectiveness of ontologies.

CNNs can be used to process textual data, such as documents, articles, or web pages, which serve as valuable sources of information for ontology generation. By treating text as a sequence of words or characters, CNNs can extract relevant features and patterns from the text data.PSO abbreviates Particle Swarm Optimization, an optimization algorithm that derives its inspiration from the collective dynamics observed in social organisms, akin to the coordinated flight of birds or the synchronized movement of fish schools. The primary objective of PSO is to unearth the optimal solution to a specified problem by simulating the cooperative conduct exhibited by a swarm of particles.

PSO can be applied in ontology generation to explore and optimize the design space of ontologies. By leveraging the swarm intelligence and optimization capabilities of PSO, ontology generation can benefit from automated exploration, optimization, and refinement of ontologies, leading to more effective and accurate representations of domain knowledge.PSO can aid in ontology alignment tasks, which involve mapping and integrating multiple ontologies. The particles can represent different candidate mappings, and the fitness function can assess the similarity and compatibility of the aligned ontologies.LSTM, which stands for Long Short-Term Memory, constitutes a distinctive breed of recurrent neural network (RNN) architecture that finds extensive application in the domain of natural language processing (NLP). This network design is meticulously crafted to apprehend intricate interdependencies across extended sequences and adeptly manage sequential data. Its prowess lies in its aptitude to adeptly retain and relinquish information over temporal horizons, thereby effectively mastering the nuances of time-evolving data.Within LSTM networks, there exist memory cells along with an array of gates that govern the intricate flow of data. The memory cells demonstrate proficiency in retaining information across extended sequences, while the gates wield their authority in meticulously orchestrating the data's trajectory. This orchestrated dance enables the network to judiciously toggle between retaining pertinent information and discreetly releasing less crucial details.

Hybrid learning approaches can be utilized, combining the power of LSTM networks with other AI techniques such as knowledge graphs, symbolic reasoning, or rule-based systems. The LSTM network can learn from the data and identify patterns, while the other AI techniques can provide additional reasoning capabilities and incorporate expert knowledge.LSTM classifiers are utilized to classify and categorize elements using neural networks while combined with methods such as STM results in a comprehensive and accurate set of ontologies

4. Results and Performance Evaluation

The dataset which was used for experimentations is an integrated dataset comprising three distinct datasets namely the Microeconomics Work With Data (2022), Eintetic (2021). Microeconomics II and Indian Economy Dataset, and the Work with Data (2023). Intermediate microeconomics. A Modern Approach Dataset. These three datasets were integrated strategically using strategic integration technique by a common annotation tool to generate annotations and based on these annotations, the reproductiveness of records of the datasets took place and for each of these annotations and the term of the dataset using customized web crawler. The documents pertaining to the terms in the dataset space were maintained in the dataset and simply large contained dataset and formalized on which experimentations were conducted.

The implementation was conducted using Python 3 as the language of choice with Google Collaboratory as the preferred IDE. Python general libraries were encompassed to conduct Natural Language Processing tasks. The WikiData was accessed through API, YAGO was accessed through SPARQL querying. CNN classifier and LSTM classifier was configured using the Keras platform. Metadata was generated using the Dspace tool

The efficacy of the introduced framework OSGAM, a model tailored for synthesizing and generating ontologies within the microeconomics realm via a fusion of hybrid learning and AI principles, is systematically assessed. The evaluation encompasses the application of standard metrics, specifically Precision-Recall Accuracy, F-Measure in its averaged form, and Fall Discovery Rates (FDR). The rationale behind adopting Precision-Recall Accuracy and F-Measure as primary metrics is rooted in their capacity to gauge the pertinence of outcomes. Additionally, the Fall Discovery Rate serves to quantitatively assess the count of false positives generated by the model. This preliminary assessment lays the foundation for the ensuing results analysis.

Table 1. Comparison of Performance of the proposed OSGAM with other approaches

Model	**Average Precision %**	**Average Recall %**	**Average Accuracy %**	**Average F-Measure %**	**FDR**
TOT	88.15	89.74	88.945	88.937	0.12
ODL	88.92	89.47	89.195	89.194	0.12
AON	90.37	92.73	91.55	91.534	0.10
Proposed OSGAM	95.09	96.84	95.965	**95.957**	0.05

From Table 1, it is indicated that the proposed OSGAM model has yielded the highest precision percentage of 95.09%, highest average recall percentage of 96.84%, highest F-measure percentage of 95.957%, lowest value of FDR of 0.05. In order to evaluate the performance of OSGAM and compare it with other prospective models, it is baselined with TOT, ODL, and AON frameworks. TOT, ODL, and AON are also ontology synthesis modeling and generation frameworks but use different techniques. TOT model has yielded 88.15% average precision, 89.74% average recall, 88.945% average accuracy, 88.937% average F-measure and finally an FDR value of 0.12. In the ODL framework, the results have been 88.92% average precision, 89.47% average recall, 89.194% average accuracy, 89.19% average F-measure, and FDR value of 0.12. The AON framework has furnished 90.37% average precision, 92.73% average recall, 91.534% F-measure, 91.55% average accuracy, and FDR value of 0.10.

Compared to all these models, the proposed OSGAM model showcases superior outcomes across various dimensions. Notably OSGAM attains the highest percentages, while simultaneously demonstrating the lowest FDR value. This remarkable performance can be attributed to the distinguishing characteristics inherent in the proposed OSGAM model, which is a semantics oriented artificially intelligent model that is based on aggregation of auxiliary knowledge in a systematic manner. The dataset itself is outgrown by identifying the relevant information terms with the help of TF-IDF and then subjecting it to topic modeling using LSTM and loading entities through community contributed, community-sourced, and community verified knowledge store repositories such as Wikidata and YAGO. Furthermore, the presence of resource description framework (RDF) helps in further contextualizing the entities through and discovering Subject and Object. Most importantly, metadata substantially augments the reservoir of knowledge infused into the localized framework, effectively bridging the cognitive divide between the existing knowledge and knowledge assimilated by the model. Further to this, the CNN and LSTM frameworks leverage the potency of deep learning models to implement robust learning techniques. In this context, the CNN framework excels in metadata classification, while the LSTM framework proficiently categorizes the documents from datasets.

Subsequently, a strong semantics is yielded through second order co-occurrence PMI and JACCARD similarity and NPMI strategy with difference and deviance measures. Based on the measure of semantic similarity and relevance, the reasoning for relevance is achieved. Particle Swarm Optimization (PSO) helps in optimizing the initial solution set to the most relevant and optimal solution set. Owing to these intelligent and strong incremental knowledge addition, loading of extensively large metadata, classifying the metadata using CNN, and classification of documents of dataset using LSTM as well as strong semantic relevance and similarity with the help of PMI, associate PMI, JACCARD similarity and strong optimization model in the form of PSO provides a robust infrastructure of hybrid learning techniques, this model ensures the proposed framework's exceptional performance surpasses that of potential baseline model, positioning it as the preeminent choice of ontology synthesis and automatic mode..

The disparity in the performance of the TOT model, which generates ontologies from tables, can be attributed to its inability to achieve knowledge-based reasoning. As a result, the TOT model's performance falls short of expectations, tables give a proper structure which means it is relational in nature so relational discovery from tables is good which means some level of semantics is given already. Constraints over concepts are yielded to give semantics, even though the semantics are not the strongest it is present. But, knowledge into the table is absent as the table only holds data as a relational DBMS. Similarly, the ODL model, despite its robust deep learning foundation, falls short in delivering anticipated outcomes due to its limitations in achieving comprehensive knowledge-based reasoning, with semantic comprehension, the amount of knowledge and the sources of knowledge which permeates into the model is quite limited and the relevance computation strategies and mechanisms are also absent in the model and henceforth the ODL model doesn't perform as expected when compared to the proposed framework. Likewise, the AON model's performance fails to meet the expectations in comparison to the proposed framework, even though it leverages NLP and statistical models for automated ontology generation from plain text. The statistical models with NLP provide stronger linguistic semantics whereas the statistical models help in relevance computation at a potential rate. The Lexico-syntacto patterns in the text corpus give some amount of semantics to it and some level of reasoning is achieved in the model. That is why it is better than the baseline models, but limited amount of knowledge, lack of regulatory mechanisms for managing a knowledge dataset, auxiliary knowledge addition is non-existent and also relevance computation is quite weak, henceforth the AON model lags compared to the proposed model.

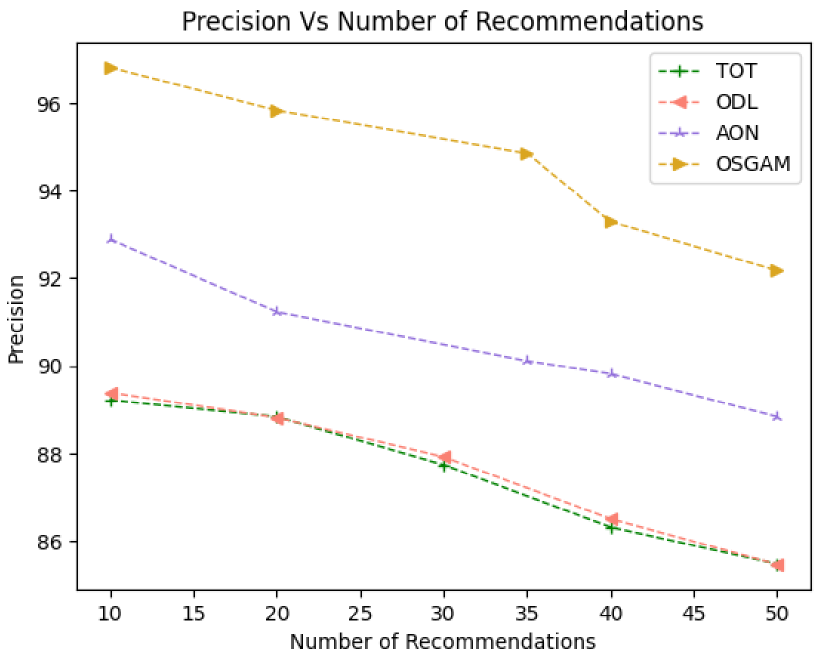

Fig. 2. Comparison of Baseline Models and Proposed Model on Precision vs Number of Recommendations

The baseline models is overcome by the proposed model and since the proposed model has a strong infrastructure of auxiliary knowledge addition in the incremental stage manner and strong ecosystem of learning through LSTM and two different CNN classifiers at certain stages and a strong relevance computation and system of semantic measures and of course, the optimization algorithm using PSO model provides a very strong infrastructure in the proposed model for semantics driven regulation of knowledge, thereby it conceives the best in class automatic model for ontology synthesis and generation.

5. Conclusion

This paper proposes a framework for knowledge synthesis model through ontologies for microeconomics as a domain of choice. Standard paradigms like TF-IDF, STM, entity harvesting through knowledge store repositories such as YAGO, Wikidata, and RDF initiating and aggregating auxiliary knowledge in the incremental manner. Metadata generation helped increase the overall knowledge density of the model at an exponential rate. Classification of the metadata using Convolutional Neural Networks helps in making it more permeable and atomic. Standard paradigms such as Second-Order co-occurrence PMI, Associate PMI, Jaccard Similarity and NPMI help in computing the semantic relatedness and thereby bolstering the semantic reasoning through semantic similarity measures. The LSTM classifier classifies the dataset which also provides a very strong learning infrastructure along with the CNN classifier which is used to classify the generated metadata at different stages of the model. The proposed framework generates a high degree of auxiliary knowledge to metadata and further classifying it and has a very strong knowledge classification infrastructure through LSTM and CNN classifier and since it has strategic schemes such as NPMI, Jaccard, for reasoning and PSO for optimization makes it the best in class model for achieving a precision of 95.09 %, F-measure percentage of 95.957 % and lowest value of FDR equals to 0.5.

References

1. Tijerino, Y. A., Embley, D. W., Lonsdale, D. W., Ding, Y., & Nagy, G. (2005). Towards ontology generation from tables. World Wide Web, 8, 261-285.
2. Wang, Y., Valipour, M., Zatarain, O. D., Gavrilova, M. L., Hussain, A., Howard, N., & Patel, S. (2017, July). Formal ontology generation by deep machine learning. In 2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) (pp. 6-15). IEEE.
3. Kumar, N., Kumar, M., & Singh, M. (2016). Automated ontology generation from a plain text using statistical and NLP techniques. International Journal of System Assurance Engineering and Management, 7, 282-293.
4. Bergamaschi, S., Castano, S., & Vincini, M. (1999). Semantic integration of semistructured and structured data sources. ACM Sigmod Record, 28(1), 54-59.
5. Biskup, J., & Embley, D. W. (2003). Extracting information from heterogeneous information sources using ontologically specified target views. Information Systems, 28(3), 169-212.
6. Castano, S., De Antonellis, V., Fugini, M. G., & Pernici, B. (1998). Conceptual schema analysis: Techniques and applications. ACM Transactions on Database Systems (TODS), 23(3), 286-333.
7. Chiang, R. H., Barron, T. M., & Storey, V. C. (1994). Reverse engineering of relational databases: Extraction of an EER model from a relational database. Data & knowledge engineering, 12(2), 107-142.
8. Poelmans, J., Ignatov, D. I., Kuznetsov, S. O., & Dedene, G. (2013). Formal concept analysis in knowledge processing: A survey on applications. Expert systems with applications, 40(16), 6538-6560.

9. Godin, R., Missaoui, R., & Alaoui, H. (1995). Incremental concept formation algorithms based on Galois (concept) lattices. Computational intelligence, 11(2), 246-267.
10. Cocchiarella, N. B. (1996). Conceptual realism as a formal ontology. In Formal ontology (pp. 27-60). Dordrecht: Springer Netherlands.
11. Allen, J. (1995). Natural language understanding. Benjamin-Cummings Publishing Co., Inc.
12. Allen, J. (1991). Natural language, knowledge representation, and logical form.