

Title: Coupled pattern learner

URL: https://en.wikipedia.org/wiki/Coupled_pattern_learner

PageID: 34042707

Categories: Category:Machine learning

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

Coupled Pattern Learner (CPL) is a machine learning algorithm which couples the semi-supervised learning of categories and relations to forestall the problem of semantic drift associated with boot-strap learning methods.

Coupled Pattern Learner

Semi-supervised learning approaches using a small number of labeled examples with many unlabeled examples are usually unreliable as they produce an internally consistent, but incorrect set of extractions. CPL solves this problem by simultaneously learning classifiers for many different categories and relations in the presence of an ontology defining constraints that couple the training of these classifiers. It was introduced by Andrew Carlson, Justin Betteridge, Estevam R. Hruschka Jr. and Tom M. Mitchell in 2009.

CPL overview

CPL is an approach to semi-supervised learning that yields more accurate results by coupling the training of many information extractors. Basic idea behind CPL is that semi-supervised training of a single type of extractor such as 'coach' is much more difficult than simultaneously training many extractors that cover a variety of inter-related entity and relation types. Using prior knowledge about the relationships between these different entities and relations CPL makes unlabeled data as a useful constraint during training. For e.g., 'coach(x)' implies 'person(x)' and 'not sport(x)'.

CPL description

Coupling of predicates

CPL primarily relies on the notion of coupling the learning of multiple functions so as to constrain the semi-supervised learning problem. CPL constrains the learned function in two ways.

Sharing among same-arity predicates according to logical relations

Relation argument type-checking

Sharing among same-arity predicates

Each predicate P in the ontology has a list of other same-arity predicates with which P is mutually exclusive. If A is mutually exclusive with predicate B, A's positive instances and patterns become negative instances and negative patterns for B. For example, if 'city', having an instance 'Boston' and a pattern 'mayor of arg1', is mutually exclusive with 'scientist', then 'Boston' and 'mayor of arg1' will become a negative instance and a negative pattern respectively for 'scientist.' Further, Some categories are declared to be a subset of another category. For e.g., 'athlete' is a subset of 'person'.

Relation argument type-checking

This is a type checking information used to couple the learning of relations and categories. For example, the arguments of the 'ceoOf' relation are declared to be of the categories 'person' and 'company'. CPL does not promote a pair of noun phrases as an instance of a relation unless the two noun phrases are classified as belonging to the correct argument types.

Algorithm description

Following is a quick summary of the CPL algorithm.

Inputs

A large corpus of Part-Of-Speech tagged sentences and an initial ontology with predefined categories, relations, mutually exclusive relationships between same-arity predicates, subset relationships between some categories, seed instances for all predicates, and seed patterns for the categories.

Candidate extraction

CPL finds new candidate instances by using newly promoted patterns to extract the noun phrases that co-occur with those patterns in the text corpus. CPL extracts,

Category Instances

Category Patterns

Relation Instances

Relation Patterns

Candidate filtering

Candidate instances and patterns are filtered to maintain high precision, and to avoid extremely specific patterns. An instance is only considered for assessment if it co-occurs with at least two promoted patterns in the text corpus, and if its co-occurrence count with all promoted patterns is at least three times greater than its co-occurrence count with negative patterns.

Candidate ranking

CPL ranks candidate instances using the number of promoted patterns that they co-occur with so that candidates that occur with more patterns are ranked higher. Patterns are ranked using an estimate of the precision of each pattern.

Candidate promotion

CPL ranks the candidates according to their assessment scores and promotes at most 100 instances and 5 patterns for each predicate. Instances and patterns are only promoted if they co-occur with at least two promoted patterns or instances, respectively.

Meta-Bootstrap Learner

Meta-Bootstrap Learner (MBL) was also proposed by the authors of CPL. Meta-Bootstrap learner couples the training of multiple extraction techniques with a multi-view constraint, which requires the extractors to agree. It makes addition of coupling constraints on top of existing extraction algorithms, while treating them as black boxes, feasible. MBL assumes that the errors made by different extraction techniques are independent. Following is a quick summary of MBL.

Subordinate algorithms used with MBL do not promote any instance on their own, they report the evidence about each candidate to MBL and MBL is responsible for promoting instances.

Applications

In their paper authors have presented results showing the potential of CPL to contribute new facts to existing repository of semantic knowledge, Freebase

See also

Co-training

Never-Ending Language Learning

Notes

References

Liu, Qihua; Xuejun Liao; Lawrence Carin (2008). "Semi-supervised multitask learning". NIPS .

Shinyama, Yusuke; Satoshi Sekine (2006). "Preemptive information extraction using unrestricted relation discovery". HLT-Naacl .

Chang, Ming-Wei; Lev-Arie Ratinov; Dan Roth (2007). "Guiding semi-supervision with constraint driven learning". ACL .

Banko, Michele; Michael J. Cafarella; Stephen Soderland; Matt Broadhead; Oren Etzioni (2007). "Open information extraction from the web". IJCAI .

Blum, Avrim; Tom Mitchell (1998). "Combining labeled and unlabeled data with co-training". Proceedings of the eleventh annual conference on Computational learning theory . pp. 92– 100. doi : 10.1145/279943.279962 . ISBN 1581130570 . S2CID 207228399 .

Riloff, Ellen; Rosie Jones (1999). "Learning dictionaries for information extraction by multi-level bootstrapping". AAAI .

Rosenfeld, Benjamin; Ronen Feldman (2007). "Using corpus statistics on entities to improve semi-supervised relation extraction from the web". ACL .

Wang, Richard C.; William W. Cohen (2008). "Iterative set expansion of named entities using the web". ICDM .