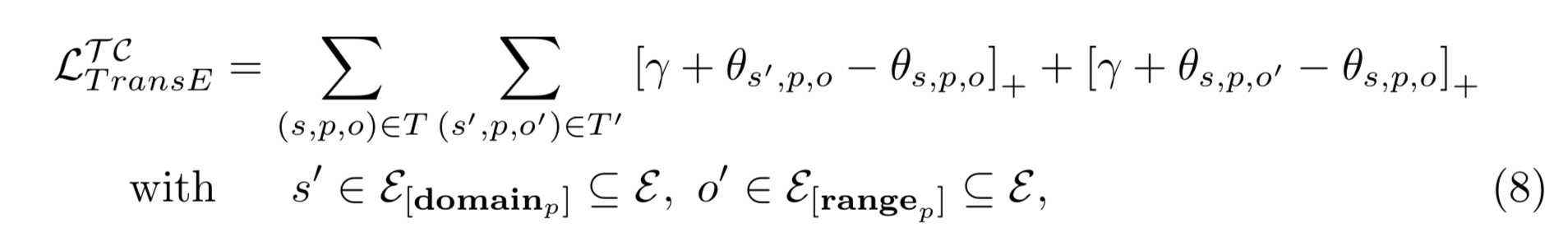
* Schema based knowledge graphs
* Integrate prior-knowledge in the form of type-constraints
* Type-constraints are not always complete or available
* Works on local closed world assumption
* Embeddings should be meaningful in low dimensional latent spaces
* Higher dimension leads to higher complexities
* Latent variable model represented as a compressed probabilistic knowledge representation
* Latent variable is opposite of observable variables – they are not directly observed but inferred through a mathematical model
* Curated prior knowledge on relation types is fed to machines e.g. marriedTo relate to only class Person
* Study impact of prior knowledge on
  + TransE
  + RESCAL: <http://www.icml-2011.org/papers/438_icmlpaper.pdf> tensor factorization bilinear model
  + Google Knowledge Vault project: <https://www.cs.ubc.ca/~murphyk/Papers/kv-kdd14.pdf> exploits non-linear interactions of latent embeddings in its neural network layers.
* Type constraints suffer from incompleteness
* Integration of LCWA and type-constraints into above three models

Prior-Knowledge on Relation-Type Semantics

* Entities in KBs are assigned to semantically predefined classes called types in hierarchical structure
* Rdf schema has rdfs:domain and rdfs:range for this purpose
* Domain covers subject entity classes and range covers object entity classes
* Example marriedTo can be used for two Persons

Type-Constrained Stochastic Gradient Descent

* in the original transE triples are not corrupted semantically
* corrupt the entities which belong to the domain of that predicate
* corrupt the entities which belong to the range of that predicate



Local Closed World Assumption

* type constraint can suffer from incompleteness and inconsistency of the data leading to fuzzy type constraints
* because of increasing fuzziness we can lose some true facts
* some triples might not have any semantic meaning
* LCWA domain and range constraints of the targeted relation type based on instance level
* Domain of a relation type-k consists of all entities that are related by the relation type-k as subject
* Range consists of all entities related as object by relation-k
* Contains undetected false triples in graph which are assumed to be true

Experimental Setup

* Explore settings
  + assume that curated type-constraints extracted from the KG’s schema available
  + explore LCWA
* enforce low dimensional latent embeddings
* combined datasets like Freebase, YAGO and DBpedia to cover diverse relation types

A screenshot of a cell phone

Description automatically generated

Evaluation Procedure

* deleting triples from the database and predicting them without using them in training
* 20%- holdout, 10%- validation, 70%- training set
* Sampled ten times negative triples for evaluation