Big Data Analysis -- Homework 2

201250182 郑义

论文名称：**Predicting Semantic Relations using Global Graph Properties**

一、Abstract 和 Introduction 翻译

Abstract

原文：Semantic graphs, such as WordNet, are resources which curate natural language on two distinguishable layers. On the local level, individual relations between synsets (semantic building blocks) such as hypernymy and meronymy enhance our understanding of the words used to express their meanings. Globally, analysis of graph-theoretic properties of the entire net sheds light on the structure of human language as a whole. In this paper, we combine global and local properties of semantic graphs through the framework of Max-Margin Markov Graph Models (M3GM), a novel extension of Exponential Random Graph Model (ERGM) that scales to large multi-relational graphs. We demonstrate how such global modeling improves performance on the local task of predicting semantic relations between synsets, yielding new state-of-the-art results on the WN18RR dataset, a challenging version of WordNet link prediction in which “easy” reciprocal cases are removed. In addition, the M3GM model identifies multi-relational motifs that are characteristic of well-formed lexical semantic ontologies.

翻译：语义图，比如 WordNet，是在两个不同的层级上区分自然语言的资源。在局部层面上，synset（语义相似构成的块）之间的个别关系，比如 hypernymy 和 meronymy 增加了我们对于表达其含义的词的理解。在全球范围内，对整个网络的图论属性的分析揭示了人类语言的整体结构。在本文中，我们通过 Max-Margin Markov Graph Model（M3GM）的框架将语义图的全局和局部属性结合起来，M3GM 是指数随机图模型（ERGM）的一种新拓展，可扩展到大型多边关系图。我么展示了这种全局建模是如何提高预测 synset 之间的语义关系这一局部关系的性能的，在 WN18RR 数据集上产生新的最先进的结果，这是 WordNet 链接预测的一个具有挑战性的版本，其中删除了“简单”的互惠情况。此外，M3GM 模型识别多相关基序，这些基序是形式良好的词法语义本体的特征。

Introduction

原文：Semantic graphs, such as WordNet(Fellbaum, 1998), encode the structural qualities of language as a representation of human knowledge. On the local level, they describe connections between specific semantic concepts, or synsets, through individual edges representing relations such as hypernymy (‘is-a’) or meronymy (‘is-part-of’); on the global level, they encode emergent regular properties in the induced relation graphs. Local properties have been subject to extensive study in recent years via the task of relation prediction, where individual edges are found based mostly on distributional methods that embed synsets and relations into a vector space (e.g. Socher et al.2013; Bordes et al., 2013; Toutanova and Chen, 2015; Neelakantan et al., 2015). In contrast, while the structural regularity and significance of global aspects of semantic graphs is wellattested (Sigman and Cecchi, 2002), global properties have rarely been used in prediction settings. In this paper, we show how global semantic graph features can facilitate in local tasks such as relation prediction.

翻译：语义图，比如 WordNet（Fellbaum， 1998），将语言的结构质量编码为人类知识的表示。在局部层面上，他们描述了具体的语义概念，或称为 synsets（语料集）之间的联系，通过代表关系的个别边，比如 hypernymy（‘is-a’） 或者 meronymy（‘is-part-of’）；在全球层面上，他们编码了诱导关系图中出现的属性规则。近年来，通过关系预测的任务对局部属性进行了广泛的研究，其中主要基于将语料集和关系嵌入到向量空间中的分布方法来发现个别边（例如 Socher 等人，2013； Bordes 等人，2013；Toutanova 和 Chen，2015；Neelakantan 等人，2015）。相比之下，虽然语义图的结构规律性和全局方面的重要性得到了很好的证明（Sigman 和 Cecchi，2002），但全局属性很少被用于预测环境中。在本文中，我们展示了全局语义图特征如何促进局部任务，如关系预测。

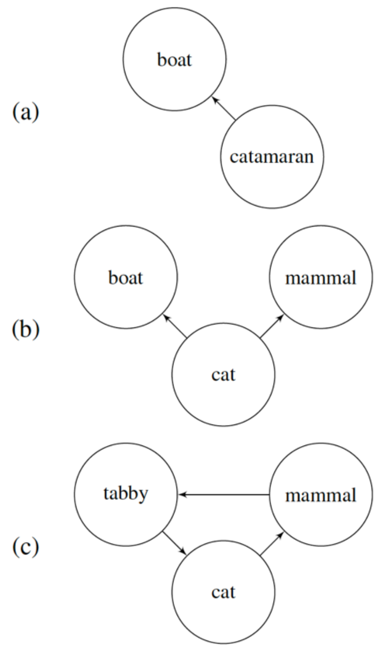

Figure 1: Probable (a) and improbable (b-c) structures in a hypothetical hypernym graph.

原文：To motivate this approach, consider the hypothetical hypernym graph fragments in Figure 1: in (a), the semantic concept (synset) ‘catamaran’ has a single hypernym, ‘boat’. This is a typical property across a standard hypernym graph. In (b), the synset ‘cat’ has two hypernyms, an unlikely event. While a local relation prediction model might mistake the relation between ‘cat’ and ‘boat’ to be plausible, for whatever reason, a high-order graph-structure-aware model should be able to discard it based on the knowledge that a synset should not have more than one hypernym. In (c), an impossible situation arises: a cycle in the hypernym graph leads each of the participating synsets to be predicted by transitivity as its own hypernym, contrary to the relation’s definition. However, a purely local model has no explicit mechanism for rejecting such an outcome.
翻译：为了激励这种方法，请考虑图 1 中假设的超义词图片段：在(a)，语义概念（synset）“双体船”有一个超子词“船”。这是标准超词图的一个典型特征。在(b)中，语义概念“猫”有两个超词，这是一个不太可能的时间。虽然局部关系预测模型可能会误认为“猫”和“船”之间的关系是可信的，但不管出于什么原因，一个高阶图结构感知模型应该能够根据一个同位素不应该有超过一个超词的知识拍抛弃它。在(c)中，一个不可能的情况出现了：超义词图中的一个循环导致每个参与的 synset 被传递性预测为它自己的超义词，这与关系的定义相反。然而，一个纯粹的局部模型没有明确的机制来拒绝这种结果。

原文：In this paper, we examine the effect of global graph properties on the link structure via the WordNet relation prediction task. Our hypothesis is that features extracted from the entire graph can help constrain local predictions to structurally sound ones (Guo et al., 2007). Such features are often manifested as aggregate counts of small subgraph structures, known as motifs, such as the number of nodes with two or more outgoing edges, or the number of cycles of length 3. Returning to the example in Figure 1, each of these features will be affected when graphs (b) and (c) are evaluated, respectively.
翻译：本文通过 WordNet 关系预测任务，研究了全局图属性对链路结构的影响。我们的假设是，从整个图中提取的特征可以帮助将局部预测限制为结构上合理的预测（Guo 等人，2007）。这些特征通常表现为小子图结构的聚合计数，称为图案，例如具有两个或多个传出边的节点数，或长度为 3 的循环数。回到图 1 中的示例，分别评估图（b）和（c）时，这些功能中的每一个都会受到影响。

原文：To estimate weights on local and global graph features, we build on the Exponential Random Graph Model (ERGM), a log-linear model over networks utilizing global graph features (Holland and

Leinhardt, 1981). In ERGMs, the likelihood of a graph is computed by exponentiating a weighted sum of the features, and then normalizing over all possible graphs. This normalization term grows exponentially in the number of nodes, and in general cannot be decomposed into smaller parts. Approximations are therefore necessary to fit ERGMs on graphs with even a few dozen nodes, and the largest known ERGMs scale only to thousands of nodes (Schmid and Desmarais, 2017). This is insufficient for WordNet, which has an order of 10^5 nodes.
翻译：为了评估局部和全局图特征的权重，我们建立了指数随机图模型（ERGM)，这是一个利用全局图特征的网络对数线性模型（Holland and Leinhardt，1981）. 在 ERGMs 中，图形的可能性是通过对特征的加权和求幂，然后对所有可能的图形进行归一化来计算的。此归一化项在节点数上呈指数级增长，并且通常不能分解为更小的部分。因此，在只有几十个节点的图上拟合 ERGM 是必要的，而已知最大的 ERGM 只能扩展到数千个节点（Schmid 和 Desmarais，2017)。这对于具有 10^5 个节点的阶数的 WordNet 是不够的。

原文：We extend the ERGM framework in several ways. First, we replace the maximum likelihood objective with a margin-based objective, which compares the observed network against alternative networks; we call the resulting model the Max-Margin Markov Graph Model (M3GM), drawing on ideas from structured prediction (Taskar et al., 2004). The gradient of this loss is approximated by importance sampling over candidate negative edges, using a local relational model as a proposal distribution. The complexity of each epoch of estimation is thus linear in the number of edges, making it possible to scale up to the 10^5 nodes in WordNet. Second, we address the multi-relational nature of semantic graphs, by incorporating a combinatorial set of labeled motifs. Finally, we link graph-level relational features with distributional information, by combining the M3GM with a dyad-level model over word sense embeddings.
翻译：我们以多种方式扩展了 ERGM 框架。首先，我们将最大似然目标替换为基于边际的目标，该目标将观察到的网络与替代网络进行比较；我们将结果模型称为最大边际马尔可夫图模型（M3GM)，借鉴了结构化预测的想法（Taskar 等人，2004)。这种损失的梯度是通过对候选负边进行重要性采样来近似的，使用局部关系模型作为建议分布。因此，每个迭代轮次估计的复杂性在边数上都是线性的，因此可以在 WordNet 中扩展到10^5个节点。其次，我们通过合并一组组合的标记主题来解决语义图的多关系性质。最后，我们将图形级关系特征与分布信息联系起来，方法是将 M3GM 与词义嵌入的二元级模型相结合。

原文：We train M3GM as a re-ranker, which we apply to a a strong local-feature baseline on the WN18RR dataset (Dettmers et al., 2018). This yields absolute improvements of 3-4 points on all commonly-used metrics. Model inspection reveals that M3GM assigns importance to features from all relations, and captures some interesting interrelational properties that lend insight into the overall structure of WordNet.

翻译：我们将 M3GM 训练为重新排名器，并将其应用于 WN18RR 数据集上的强局部特征基线（Dettmers 等人，2018）。这在所有常用指标上产生了 3-4 个点的绝对改进。模型检查显示，M3GM 对所有关系中的特征都赋予了重要性，并捕获了一些有趣的关系间属性，这些属性可以深入了解 WordNet 的整体结构。

二、问题描述

本文主要是为了解决之前的链接预测模型只从局部来对一个关系进行分析，这可能导致分析结果违反关系的定义，本文从 ERGM 提出一种新的 M3GM 来从局部和全局对链接关系进行分析，构建模型来预测关系。

三、输入、输出、模型算法描述

输入：关系预测数据集 WN18。包含从 WordNet 3.0 中提取的约 41,000 个同义词集的 18 个关系。

输出：最终输出的为一个经过多轮训练之后的 Model，可以用来进行预测

模型算法描述：考虑一个图 G = (V,E)，V 为顶点，E 为有向边，ERGM 评分函数定义了 G|V ∣ 上的概率，即具有 |V ∣ 的所有图的集合。这个概率被定义为一个对数线性函数：

$$P_{\text{ERGM}}(G) \propto \psi_{\text{ERGM}}(G) = \exp\left(\boldsymbol{\theta}^T \mathbf{f}(G)\right)$$

其中 f 是一个特征函数，从图到特征计数向量。 特征通常是图案的数量——小的子图结构——如简介中所述。 向量是要估计的参数。 在本节中，我们将讨论我们将该模型应用于语义图领域，利用它们的特殊属性。 语义图由多种关系类型组成，特征空间需要容纳这些关系类型； 它们的节点是与复杂解释相关的语言结构（语义概念），可以通过将它们在 Rd 中的嵌入合并到新的评分模型中来使图形表示受益。 然后，作者展示了 M3GM 框架，以对新模型执行可靠和高效的参数估计。

以 ERGM 为基础，提出的新特点：在经典的 ERGM 应用领域，如社交媒体或生物网络，节点往往具有很少的内在区别，或者至少在应用模型之前可以提取的有意义的内在信息很少。 然而，在语义图中，节点表示同义词集，这些同义词集与对预测图结构很有价值的信息相关联，并且可以使用无监督技术进行近似，例如基于大量可用数据嵌入到公共 d 维向量空间中。 因此，我们从 eq 修改了传统的评分函数。 通过引入特定于关系的关联运算符来包含特定于节点的信息：

erator $\mathcal{A}^{(r)} : V \times V \to \mathbb{R}$:

$$\begin{aligned} &\psi_{\text{ERGM+}}(G) = \\ &= \exp\left(\boldsymbol{\theta}^T \mathbf{f}(G) + \sum_{r \in \mathcal{R}} \sum_{(s,t) \in E_r} \mathcal{A}^{(r)}(s,t)\right). \end{aligned} \tag{2}$$

The association operator generalizes various models from the relation prediction literature:

TransE (Bordes et al., 2013) embeds each relation r into a vector in the shared space, representing a 'difference' between sources and targets, to compute the association score under a translational objective,

$$\mathcal{A}^{(r)}_{\text{TRANSE}}(s,t) = -\|\boldsymbol{e}_s + \boldsymbol{e}_r - \boldsymbol{e}_t\|.$$

BiLin (Nickel et al., 2011) embeds relations into full-rank matrices, computing the score by a bilinear multiplication,

$$\mathcal{A}^{(r)}_{\text{BILIN}}(s,t) = \boldsymbol{e}_s^T \mathbf{W}_r \boldsymbol{e}_t.$$

DistMult (Yang et al., 2014) is a special case of **BiLin** where the relation matrices are diagonal, reducing the computation to a ternary dot product,

$$\mathcal{A}^{(r)}_{\text{DISTMULT}}(s,t) = \langle \boldsymbol{e}_s, \boldsymbol{e}_r, \boldsymbol{e}_t \rangle = \sum_{i=1}^{d} e_{s_i}\, e_{r_i}\, e_{t_i}.$$

四、评价指标及计算公式

表 1 展示了开发集上的结果。 第 1-3 行描述了使用平均 FastText 嵌入初始化的本地模型的结果，表明 TRANSE 实现了 MRR 和顶级命中方面的最佳性能。平均排名与其他指标不一致； 这是一个可解释的权衡，因为 BILIN 和 DISTMULT 都对相关的同义词集嵌入具有固有的偏好，在关系嵌入完全关闭的情况下提供更强的回退，但与从相关误报中分离强案例的自由度相比。

1. *Effect of global score.*
2. *Synset embedding initialization.*

	System	MR	MRR	H@10	H@1
	RULE	13396	35.26	35.27	35.23
1	DISTMULT	1111	43.29	50.73	39.67
2	BILIN	**738**	45.36	52.93	41.37
3	TRANSE	2231	**46.07**	**55.65**	**41.41**
4	M3GM	2231	47.94	**57.72**	43.26
5	M3GM$_{\alpha_r}$	2231	**48.30**	57.59	**43.78**

Table 1: Results on development set (all metrics except MR are x100). M3GM lines use TRANSE as their association model. In M3GM$_{\alpha_r}$, the graph component is tuned post-hoc against the local component per relation.

System	MR	MRR	H@10	H@1
RULE	13396	35.26	35.26	35.26
COMPLEX†	5261	44	51	41
CONVE†	5277	46	48	39
CONVKB†	2554	24.8	52.5	
TRANSE	2195	46.59	55.55	42.26
M3GM$_{\alpha_r}$	**2193**	**49.83**	**59.02**	**45.37**

Table 2: Main results on test set. † These models were not re-implemented, and are reported as in Nguyen et al. (2018) and in Dettmers et al. (2018).

Positive	
1	$s \xrightarrow{member\_meronym} t$
2	$s \xrightarrow{has\_part} t$
3	$s \xrightarrow{hypernym} t \xrightarrow{derivationally\_related\_form} u$
Negative	
4	$s \xrightarrow{hypernym} t$
5	$s \xleftrightarrow{hypernym} t$
6	$s \xrightarrow{member\_meronym} t \xrightarrow{instance\_hypernym} u$
7	$s_1 \xrightarrow{has\_part} t \xleftarrow{verb\_group} s_2$

Table 3: Select heavyweight features (motifs) following best dev set training using M3GM. Circled nodes count towards the motif.

Source	Relation	Correct target	Outranking local target(s)
indian lettuce	*hypernym*	herb	garden lettuce
austria	*has_part*	vienna	germany, hungary, france, european union

Table 4: Successful M3GM re-ranking examples.

Relation r	α_r	Relation r	α_r
mem. of domain usage	0.78	hypernym	0.64
mem. of domain region	0.77	domain topic of	0.38
member meronym	0.67	has part	0.33
instance hypernym	0.65		

Table 5: Graph score weights found for relations on the dev set. Zero means graph score is not considered at all for this relation, one means only it is considered.

needed Mimick initialization.

五、对比方法及引用出处

ERGM，https://eehh-stanford.github.io/SNA-workshop/ergm-intro.html。

六、结果

通过 20 批次的训练，的到了 M3GM，能够很好的预测链接关系，可看附件。

数据集链接：

1. WN18:https://github.com/villmow/datasets_knowledge_embedding/tree/master/WN18RR
2. FastText: https://fasttext.cc/