SemEval-2013 Task 3: Spatial Role Labeling

Oleksandr Kolomiyets[†], **Parisa Kordjamshidi**[†],
Steven Bethard[‡] and **Marie-Francine Moens**[†]
[†]KU Leuven, Celestijnenlaan 200A, Heverlee 3001, Belgium
[‡]University of Colorado, Campus Box 594 Boulder, Colorado, USA

Abstract

Many NLP applications require information about locations of objects referenced in text, or relations between them in space. For example, the phrase *a book on the desk* contains information about the location of the object *book*, as *trajector*, with respect to another object *desk*, as *landmark*. Spatial Role Labeling (SpRL) is an evaluation task in the information extraction domain which sets a goal to automatically process text and identify objects of spatial scenes and relations between them. This paper describes the task in Semantic Evaluations 2013, annotation schema, corpora, participants, methods and results obtained by the participants.

1 Introduction

Spatial Role Labeling at SemEval-2013 is the second iteration of the task, which was initially introduced at SemEval-2012 (Kordjamshidi et al., 2012a). The second iteration extends the previous work with an additional training corpus, which contains besides "static" spatial relations, annotated motions. Motion detection is a novel task for annotating trajectors (objects, which are moving), landmarks (spatial context in which the motion is performed), motion indicators (lexical triggers which signals trajector's motion), paths (a path along which the motion is performed), directions (absolute or relative directions of trajector's motion) and distances (a distance as a product of motion). For annotating motions the existing annotation scheme has been adapted with additional markables which are, all together, described below.

2 Spatial Annotation Schema

In this Section we describe the annotation format of spatial markables in text, and annotation guidelines for the annotators.

2.1 Spatial Annotation Format

Building upon the previous work, we used the notions of trajectors, landmarks and spatial indicators as introduced by Kordjamshidi *et al.* (2010). In addition, we further expanded the set of spatial roles labels with motion indicators, paths, directions and distances to capture fine-grained spatial semantics of *static* spatial relations (as the ones which do not involve motions), and to accommodate *dynamic* spatial relations (the ones which do involve motions).

2.1.1 Static Spatial Relations and their Roles

Static spatial relations are defined as relations between still objects, whereas one object plays a central role in the spatial scene, which is called *trajector*, and the second one plays a secondary role, and it is called *landmark*. In language, a spatial relation between two objects is usually implemented by a preposition (*in*, *on*, *at*, etc.) or a prepositional phrase (*on top of*, *inside of*, etc.).

A static spatial relation is defined as a tuple that contains a trajector, a landmark and a spatial indicator. In the annotation schema, these annotations are defined as follows:

Trajector: Trajector is a spatial role label assigned to a word or a phrase that denotes a central object of a spatial scene. For example:

- [$_{Trajector}$ *a lake*] *in the forest*

255

- [$_{Trajector}$ *a flag*] *on top of the building*

Landmark: Landmark is a spatial role label assigned to a word or a phrase that denotes a secondary object of a spatial scene, to which a possible spatial relation (as between two objects in space) can be established. For example:

- *a lake in* [$_{Landmark}$ *the forest*]

- *a flag on top of* [$_{Landmark}$ *the building*]

Spatial Indicator: Spatial Indicator is a spatial role label assigned to a word or a phrase that signals a spatial relation between objects (trajectors and landmarks) of a spatial scene. For example:

- *a lake* [$_{Sp\ indicator}$ *in*] *the forest*

- *a flag*] [$_{Sp\ indicator}$ *on top of*] *the building*

Spatial Relation: Spatial Relation is a relation that holds between spatial markables in text as, e.g., between a trajector and a landmark and triggered by a spatial indicator. In spatial information theory the relations and properties are usually grouped into the domains of topological, directional, and distance relations and also shape (Stock, 1998). Three semantic classes for spatial relations were proposed:

- **Region.** This type refers to a region of space which is always defined in relation to a landmark, e.g., the interior or exterior. For example:

  a lake in the forest $\Longrightarrow$ $\langle$Region, [$_{Sp\ indicator}$ *in*], [$_{Trajector}$ *a lake*], [$_{Landmark}$ *the forest*]$\rangle$

- **Direction.** This relation type denotes a direction along the axes provided by the different frames of reference, in case the trajector of motion is not characterized in terms of its relation to the region of a landmark. For example:

  a flag on top of the building $\Longrightarrow$ $\langle$Direction, [$_{Sp\ indicator}$ *on top of*], [$_{Trajector}$ *a flag*], [$_{Landmark}$ *the building*]$\rangle$

- **Distance.** Type *Distance* states information about the spatial distance of the objects and could be a qualitative expression, such as *close*, far or quantitative, such as *12 km*. For example:

  the kids are close to the blackboard $\Longrightarrow$ $\langle$Distance, [$_{Distance}$ *close*], [$_{Trajector}$ *the kids*], [$_{Landmark}$ *the blackboard*]$\rangle$

2.1.2 Dynamic Spatial Relations

In addition to static spatial relations and their roles, SpRL-2013 introduces new spatial roles to capture *dynamic* spatial relations which involve motions. Let us demonstrate this with the following example:

(1)   *In Brazil coming from the North-East I stepped into the small forest and followed down a dried creek.*

The text above describes a motion, and the reader can identify a number of concepts which are peculiar for motions: there is an object whose location is changing, the motion is performed in a specific spatial context, with a specific direction, and with a number of locations related to the object's motion.

There has been an enormous effort in formalizing and annotating motions in natural language. While annotating motions was out of scope for the previous SpRL task and SpatialML (Mani et al., 2010), the most recent work on the Dynamic Interval Temporal Logic (DITL) (Pustejovsky and Moszkowicz, 2011) presents a framework for modeling motions as a change of state, which adapts linguistic background considering path constructions and *manner-of-motion* constructions. On this basis the Spatiotemporal Markup Language (STML) has been introduced for annotating motions in natural language. In STML, a motion is treated as a change of location over time, while differentiating between a number of spatial configurations along the path. Being well-defined for the formal representations of motion and reasoning, in which representations either take explicit reference to temporal frames or reify a spatial object for a path, all the previous work seems to be difficult to apply in practice when annotating motions in natural language. It can be attributed to possible vague descriptions of path in natural language when neither clear temporal event ordering, nor distinction between the start, end or intermediate path point can be made.

In SpRL-2013, we simplify the previously introduced notion of path in order to provide practical motion annotations. For dynamic spatial relations we introduce the following roles:

256

Trajector: Trajector is a spatial role label assigned to a word or a phrase which denotes an object which moves, starts, interrupts, resumes a motion, or is forcibly involved in a motion. For example:

- ... *coming from the North-East* [$_{Trajector}$ *I*] *stepped into* ...

Motion Indicator: Motion indicator is a spatial role label assigned to a word or a phrase which signals a motion of the trajector along a path. In Example (1), a number of motion indicators can be identified:

- ... [$_{Motion}$ *coming*] *from the North-East I* [$_{Motion}$ *stepped into*] ... *and* [$_{Motion}$ *followed down*] ...

Path: Path is a spatial role label assigned to a word or phrase that denotes the path of the motion as the trajector is moving along, starting in, arriving in or traversing it. In SpRL-2013, as opposite to STML, the notion of path does not have the temporal dimension, thus whenever the motion is performed along a path, for which either a start, an intermediate, an end path point, or an entire path can be identified in text, they are labeled as path. In Example (1), a number of path labels can be identified:

- ... *coming* [$_{Path}$ *from the North-East*] *I stepped into* [$_{Path}$ *the small forest*] *and followed down* [$_{Path}$ *a dried creek*].

Landmark: The notion of path should not be confused with landmarks. For spatial annotations, landmark has been introduced as a spatial role label for a secondary object of the spatial scene. Being of great importance for static spatial relations, in dynamic spatial relations, landmarks are used to capture a spatial context of a motion as for example:

- *In* [$_{Landmark}$ *Brazil*] *coming from the North-East* ...

Distance: In contrast to the previous SpRL annotation standard, in which distances and directions have been uniformly treated as signals, in SpRL-2013 if the motion is performed for a certain distance, and such a distance is mentioned in text, the corresponding textual span is labeled as distance.

Distance is a spatial role label assigned to a word or a phrase that denotes an absolute or relative distance of motion, or the distance between a trajector and a landmark in case of a static spatial scene. For example:

- [$_{Distance}$ *25 km*]

- [$_{Distance}$ *about 100 m*]

- [$_{Distance}$ *not far away*]

- [$_{Distance}$ *25 min by car*]

Direction: Additionally, if the motion is performed in a certain (absolute or relative) direction, and such a direction is mentioned in text, the corresponding textual span is annotated as direction. Direction is a spatial role label assigned to a word or a phrase that denotes an absolute or relative direction of motion, or a spatial arrangement between a trajector and a landmark. For example:

- [$_{Direction}$ *the North-West*]

- [$_{Direction}$ *northwards*]

- [$_{Direction}$ *west*]

- [$_{Direction}$ *the left-hand side*]

Spatial Relation: Similarly to static spatial relations, dynamic spatial relations are annotated by relations that hold between a number of spatial roles. The major difference to static spatial relations is the mandatory motion indicator[1]. For example:

- *In Brazil coming from the North-East I* ... ⟹ ⟨Direction, [$_{Sp\ indicator}$ *In*], [$_{Trajector}$ *I*], [$_{Landmark}$ *Brazil*], [$_{Motion}$ *coming*],[$_{Path}$ *from the North-East*]⟩

- ... *I stepped into the small forest and* ... ⟹ ⟨Direction, [$_{Trajector}$ *I*], [$_{Motion}$ *stepped into*],[$_{Path}$ *the small forest*]⟩

- ... *I* [...] *and followed down a dried creek.* ⟹ ⟨Direction, [$_{Trajector}$ *I*], [$_{Motion}$ *followed down*],[$_{Path}$ *a dried creek*]⟩

---

[1]All dynamic spatial relations were annotated with type *Direction*.

257

Corpus		Files	Sent.	TR	LM	SI	MI	Path	Dir	Dis	Relation
IAPR TC-12	Training	1	600	716	661	670	-	-	-	-	765
	Evaluation	1	613	872	743	796	-	-	-	-	940
Confluence	Training	95	1422	1701	1037	879	1039	945	223	307	2105
Project	Evaluation	22	367	497	316	247	305	240	37	87	598

Table 1: Corpus statistics for SpRL-2013 with respect to annotated spatial roles (trajectors (TR), landmarks (LM), spatial indicators (SI), motion indicators (MI), paths (Path), directions (Dir) and distances (Dis)) and spatial relations.

3 Corpora

The data for the shared task comprises two different corpora.

3.1 IAPR TC-12 Image Benchmark Corpus

The first corpus is a subset of the IAPR TC-12 image benchmark corpus (Grubinger et al., 2006). It contains 613 text files that include 1213 sentences in total, and represents an extension of the dataset previously used in (Kordjamshidi et al., 2011). The original corpus was available free of charge and without copyright restrictions. The corpus contains images taken by tourists with descriptions in different languages. The texts describe objects, and their absolute and relative positions in the image. This makes the corpus a rich resource for spatial information, however, the descriptions are not always limited to spatial information. Therefore, they are less domain-specific and contain free explanations about the images. For training we released 600 sentences (about 50% of the corpus), and used remaining 613 sentences for evaluations.

3.2 Confluence Project Corpus

The second corpus comes from the Confluence project that targets the description of locations situated at each of the latitude and longitude integer degree intersection in the world. This corpus contains user-generated content produced by, sometimes, non-native English speakers. We gathered the content by keeping the original orthography and formating. In addition, we stored the URLs of the descriptions and extracted the coordinates of the described confluence point, which might be interesting for further research. In total, the entire corpus contains 117 files with 1789 sentences (about 40,000 tokens). For training we released 95 annotated files with 1422 sentences, 2105 annotated relations in to-

tal. For evaluation we used 22 annotated files with 367 sentences. The statistics on both corpora are provided in Table 1.

3.3 Data Format

One important change to the data was made in SpRL-2013. In contrast to SpRL-2012, where spatial roles were annotated over "head words" whose indexes were part of unique identifiers, in SpRL-2013 we switched to span-based annotations. Moreover, in order to provide a single data format for the task, we transformed SpRL-2012 data into span-based annotations, in course of which, we identified a number of annotation errors and made further improvements for about 50 annotations.

For annotating the Confluence Project corpus we used a freely available annotation tool MAE created by Amber Stubbs (Stubbs, 2011). The resulting data format uses the same annotation tags as in SpRL-2012, but each role annotation refers to a character offset in the original text[2]. Spatial relations are composed of references to annotations by their unique identifiers. Similarly to SpRL-2012, we allowed annotators to provide non-consuming annotations, where entity mentions, for which spatial roles can be identified, are omitted in text but necessary for a spatial relation triggered by either a spatial indicator or a motion indicator. Two spatial roles are eligible for non-consuming annotations: trajectors and landmarks.

4 Tasks Descriptions

For the sake of consistency with SpRL-2012, in SpRL-2013 we proposed the following tasks:

---

[2]Due to paper length constraints we omit the BNF specifications for spatial roles and relations. For further data format information we refer the reader to the task description web page: www.cs.york.ac.uk/semeval-2013/task3/

- Task A: Identification of markable spans for **three types** of spatial annotations such as trajector, landmark and spatial indicator.

- Task B: Identification of tuples (triplets) that connect trajectors, landmarks and spatial indicators identified in Task A into spatial relations. That is, identification of spatial relations with **three markables** connected, and **without** semantic relation classification.

- Task C: Identification of markable spans for **all** spatial annotations such as trajector, landmark, spatial indicator, motion indicator, path, direction and distance.

- Task D: Identification of *n-tuples* that connect spatial markables identified in Task C into spatial relations. That is, identification of spatial relations **with as many participating markables as possible**, and **without** semantic relation classification.

- Task E: **Semantic classification** of spatial relations identified in Task D.

5 Evaluation Criteria and Metrics

System outputs were evaluated against the gold annotations, which had to conform to the role's Backus-Naur form. For Tasks A and C, the system annotations are spatial roles: spans of text associated with spatial role types. A system annotation of a role is considered correct if it has a minimal overlap of one character with a gold annotation and matches the role type of the gold annotation. For Tasks B and D, the system annotations are spatial relation tuples (of length 3 in task B, of length 3 to 5 in Task D) of references to markable annotations. A system annotation of a spatial relation tuple is considered correct if it is of the same length as the gold annotation, and if each spatial role in the system tuple matches each role in the gold tuple. A spatial role estimated by a system is considered correct if it matches a gold reference when having the same character offsets and markable types (strict evaluation settings). In addition we introduced relaxed evaluation settings, in which a minimal overlap of one character between a system and a gold markable references is required for a positive match under condition that the roles

match. For Task E, the system annotations are spatial relation tuples of length 3 to 5, along with relation type labels. A system annotation of a spatial relation is considered correct if the spatial relation tuple is correct under the evaluation of Task D and the relation type of the system relation is the same as the relation type of the gold relation.

Systems were evaluated for each of the tasks in terms of precision (*P*), recall (*R*) and *F1*-score which are defined as follows:

$$Precision = \frac{tp}{tp + fp} \qquad (1)$$

$$Recall = \frac{tp}{tp + fn} \qquad (2)$$

where tp is the number of true positives (the number of instances that are correctly found), fp is the number of false positives (number of instances that are predicted by the system but not a true instance), and fn is the number of false negatives (missing results).

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad (3)$$

6 System Description and Evaluation Results

UNITOR. The UNITOR-HMM-TK system addressed Tasks A,B and C (Bastianelli et al., 2013).

In Tasks A and C, roles are labeled by a sequence-based classifier: each word in a sentence is classified with respect to the possible spatial roles. An approach based on the SVM-HMM learning algorithm, formulated in (Tsochantaridis et al., 2006), was used. It is in line with other methods based on sequence-based classifier for Spatial Role Labeling, such as Conditional Random Fields (Kordjamshidi et al., 2011), and the same SVM-HMM learning algorithm (Kordjamshidi et al., 2012b). UNITOR's labeling approach has been inspired by the work in (Croce et al., 2012), where an SVM-HMM learning algorithm has been applied to the classical FrameNet-based Semantic Role Labeling. The main contribution of the proposed approach is the adoption of shallow grammatical features instead of the full syntax of the sentence, in order to avoid over-fitting on the training data. Moreover, lexical information has been generalized through the use

Run	Task	Evaluation	Label	P	R	F_1-score
UNITOR.Run1.1	Task A	relaxed	TR	0.684	0.681	0.682
			LM	0.741	0.835	0.785
			SI	0.967	0.889	0.926
	Task B	relaxed	Relation	0.551	0.391	0.458
		strict	Relation	0.431	0.306	0.358
UNITOR.Run1.2	Task A	relaxed	TR	0.682	0.493	0.572
			LM	0.801	0.560	0.659
			SI	0.968	0.585	0.729
	Task B	relaxed	Relation	0.551	0.391	0.458
		strict	Relation	0.431	0.306	0.358
UNITOR.Run2.1	Task A	relaxed	TR	0.565	0.317	0.406
			LM	0.661	0.476	0.554
			SI	0.612	0.481	0.538
	Task C	relaxed	TR	0.565	0.317	0.406
			LM	0.662	0.476	0.554
			SI	0.609	0.479	0.536
			MI	0.892	0.294	0.443
			Path	0.775	0.295	0.427
			Dir	0.312	0.229	0.264
			Dis	0.946	0.331	0.490

Table 2: Results of UNITOR for SpRL-2013 tasks (Task A, B and C).

of Word Space – a Distributional Model of Lexical Semantics derived from the unsupevised analysis of an unlabeled large-scale corpus (Sahlgren, 2006). Similarly to the approaches demonstrated in SpRL-2012, the proposed approach first classifies spatial and motion indicators, then, using these outcomes further spatial roles are determined. For classifying indicators, the classifier makes use of lexical and grammatical features like lemmas, part-of-speech tags and lexical context representations. The remaining spatial roles are estimated by another classifier additionally employing the lemma of the indicator, distance and relative position to the indicator, and the number of tokens composing the indicator as features.

In Task B, all roles found in a sentence for Task A are combined to generate candidate relations, which are verified by a Support Vector Machine (SVM) classifier. As the entire sentence is informative to determine the proper conjunction of all roles, a Smoothed Partial Tree Kernel (SPTK) within the classifier that enhances both syntactic and lexical information of the examples was applied (Croce et al.,

2011). This is a convolution kernel that measures the similarity between syntactic structures, which are partially similar and whose nodes can be different, but are, nevertheless, semantically related. Each example is represented as a tree-structure which is directly derived from the sentence dependency parse, and thus allows for avoiding manual feature engineering as in contrast to the work of Roberts and Harabagiu (2012). In the end, the similarity score between lexical nodes is measured by the Word Space model.

UNITOR submitted two runs for the IAPR TC-12 Image benchmark corpus (we refer to them as to UNITOR.Run1.1 and UNITOR.Run1.2) and one run for the Confluence Project corpus (UNITOR.Run2.1), based on the models individually trained on the different corpora. The difference between UNITOR.Run1.1 and UNITOR.Run1.2 is that for UNITOR.Run1.1 the results are obtained for all spatial roles (also the ones that have no spatial relation), and UNITOR.Run1.2 only provided the roles for which also spatial relations were identified. The results are presented in Table 2.

Although, not directly comparable to the results in SpRL-2012, one may observe some common trends. First, similarly to the previous findings, the performance for recognition of landmarks and spatial indicators (Task A) on the IAPR TC-12 Image benchmark corpus is better than trajectors (F_1-scores of 0.785, 0.926 and 0.682 respectively), and spatial indicators is the "easiest" spatial role to recognize (F_1-score of 0.926).

In contrast, spatial role labeling on the Confluence Project corpus performs worse than on the IAPR TC-12 Image benchmark corpus (with F_1-scores of 0.406, 0.538 and 0.554 for trajectors, spatial indicators and landmarks respectively). Interestingly, the performance for landmarks is generally higher than for trajectors, which is in line with previous findings in SpRL-2012. The performance drop on the new corpus can be attributed to more complex text and descriptions, whereas multiple roles can be identified for the same span (for example, a path which spans over trajectors, landmarks and spatial indicators). For the new spatial roles of motion indicators, paths, directions and distances, the performance levels are overall higher than for trajectors with an exception of directions. Yet, the precision levels for new roles is much higher than the recall (0.892 vs. 0.294 for motion indicators, 0.775 vs. 0.295 for paths and 0.946 vs. 0.331 for distances). Directions turned out to be the most difficult role to classify (0.312, 0.229 and 0.264 for P, R and F_1-score respectively).

7 Conclusion

In this paper we described an evaluation task on Spatial Role Labeling in the context of Semantic Evaluations 2013. The task sets a goal to automatically process text and identify objects of spatial scenes and relations between them. Building largely upon the previous evaluation campaign, SpRL-2012, in SpRL-2013 we introduced additional spatial roles and relations for capturing motions in text. In addition, a new annotated corpus for spatial roles (including annotated motions) was produced and released to the participants. It comprises a set of 117 files with about 40,000 tokens in total.

With the registered number of 10 participants and the final number of submissions (only one) we can conclude that spatial role labeling is an interesting task within the research community, however sometimes underestimated in its complexity. Our further steps in promoting spatial role labeling will be a detailed description of the annotation scheme and annotation guidelines, analysis of the corpora and obtained results.

References

Emanuele Bastianelli, Danilo Croce, Roberto Basili, and Daniele Nardi. 2013. UNITOR-HMM-TK: Structured Kernel-based learning for Spatial Role Labeling. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Association for Computational Linguistics.

Danilo Croce, Alessandro Moschitti, and Roberto Basili. 2011. Structured Lexical Similarity via Convolution Kernels on Dependency Trees. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1034–1046. Association for Computational Linguistics.

Danilo Croce, Giuseppe Castellucci, and Emanuele Bastianelli. 2012. Structured Learning for Semantic Role Labeling. *Intelligenza Artificiale*, 6(2):163–176.

Michael Grubinger, Paul Clough, Henning Müller, and Thomas Deselaers. 2006. The IAPR TC-12 Benchmark: A New Evaluation Resource for Visual Information Systems. In *International Workshop OntoImage*, pages 13–23.

Parisa Kordjamshidi, Marie-Francine Moens, and Martijn van Otterlo. 2010. Spatial Role Labeling: Task Definition and Annotation Scheme. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 413–420.

Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens. 2011. Spatial Role Labeling: Towards Extraction of Spatial Relations from Natural Language. *ACM Transactions on Speech and Language Processing (TSLP)*, 8(3):4.

Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2012a. Semeval-2012 Task 3: Spatial Role Labeling. In *Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 365–373. Association for Computational Linguistics.

Parisa Kordjamshidi, Paolo Frasconi, Martijn Van Otterlo, Marie-Francine Moens, and Luc De Raedt.

2012b. Relational Learning for Spatial Relation Extraction from Natural Language. In *Inductive Logic Programming*, pages 204–220. Springer.

Inderjeet Mani, Christy Doran, Dave Harris, Janet Hitzeman, Rob Quimby, Justin Richer, Ben Wellner, Scott Mardis, and Seamus Clancy. 2010. SpatialML: Annotation Scheme, Resources, and Evaluation. *Language Resources and Evaluation*, 44(3):263–280.

James Pustejovsky and Jessica L Moszkowicz. 2011. The Qualitative Spatial Dynamics of Motion in Language. *Spatial Cognition & Computation*, 11(1):15–44.

Kirk Roberts and Sanda M Harabagiu. 2012. UTD-SpRL: A Joint Approach to Spatial Role Labeling. In *Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 419–424. Association for Computational Linguistics.

Magnus Sahlgren. 2006. *The Word-space Model*. Ph.D. thesis, Stockholm University.

Oliviero Stock. 1998. *Spatial and Temporal Reasoning*. Springer-Verlag New York Incorporated.

Amber Stubbs. 2011. MAE and MAI: Lightweight Annotation and Adjudication Tools. In *Proceedings of the 5th Linguistic Annotation Workshop*, LAW V '11, pages 129–133, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, Yasemin Altun, and Yoram Singer. 2006. Large Margin Methods for Structured and Interdependent Output Variables. *Journal of Machine Learning Research*, 6(2):1453.