Neural-Symbolic Models in AI

Shalom Lappin

Queen Mary University of London, University of Gothenburg, and King's College London

CLASP Seminar

February 27, 2025

Outline

Some AI History

Injective Hybrid Models

Federative Hybrid Models

Conclusions and Future Work

The Early Years of AI

- In the early years of AI both neural networks and symbolic systems were unable to go beyond small scale models, which had to be adjusted, often by handcrafted extensions, to new cases.
- This was, in large measure, the result of the hardware limitations of the time, and the absence of digitalised data for training and testing.
- Feed forward neural networks lacked memory for tracking long distance dependency relations in input data.
- Symbolic systems did not incorporate learning procedures, and so their rules had to be devised by hand.
- Minsky (1991) suggested that hybrid systems, combining neural networks for lower level perceptual classification and symbolic components for reasoning, were needed for progress in AI.

The Early Years of AI

- In the early years of AI both neural networks and symbolic systems were unable to go beyond small scale models, which had to be adjusted, often by handcrafted extensions, to new cases.
- This was, in large measure, the result of the hardware limitations of the time, and the absence of digitalised data for training and testing.
- Feed forward neural networks lacked memory for tracking long distance dependency relations in input data.
- Symbolic systems did not incorporate learning procedures, and so their rules had to be devised by hand.
- Minsky (1991) suggested that hybrid systems, combining neural networks for lower level perceptual classification and symbolic components for reasoning, were needed for progress in AI.

The Early Years of AI

- In the early years of AI both neural networks and symbolic systems were unable to go beyond small scale models, which had to be adjusted, often by handcrafted extensions, to new cases.
- This was, in large measure, the result of the hardware limitations of the time, and the absence of digitalised data for training and testing.
- Feed forward neural networks lacked memory for tracking long distance dependency relations in input data.
- Symbolic systems did not incorporate learning procedures, and so their rules had to be devised by hand.
- Minsky (1991) suggested that hybrid systems, combining neural networks for lower level perceptual classification and symbolic components for reasoning, were needed for progress in AI.

The Early Years of AI

- In the early years of AI both neural networks and symbolic systems were unable to go beyond small scale models, which had to be adjusted, often by handcrafted extensions, to new cases.
- This was, in large measure, the result of the hardware limitations of the time, and the absence of digitalised data for training and testing.
- Feed forward neural networks lacked memory for tracking long distance dependency relations in input data.
- Symbolic systems did not incorporate learning procedures, and so their rules had to be devised by hand.
- Minsky (1991) suggested that hybrid systems, combining neural networks for lower level perceptual classification and symbolic components for reasoning, were needed for progress in AI.

The Early Years of AI

- In the early years of AI both neural networks and symbolic systems were unable to go beyond small scale models, which had to be adjusted, often by handcrafted extensions, to new cases.
- This was, in large measure, the result of the hardware limitations of the time, and the absence of digitalised data for training and testing.
- Feed forward neural networks lacked memory for tracking long distance dependency relations in input data.
- Symbolic systems did not incorporate learning procedures, and so their rules had to be devised by hand.
- Minsky (1991) suggested that hybrid systems, combining neural networks for lower level perceptual classification and symbolic components for reasoning, were needed for progress in AI.

The Deep Learning Revolution

- In the past three decades the emergence of powerful hardware (GPUs), the abundance of online data, and radical innovations in the architecture of neural networks, have produced the deep learning revolution.
- Transformers, which drive Large Language Models, consist entirely of blocks of attention heads.
- These are trained independently of each other, and they can identify fine grained patterns in data across distinct modalities (text, visual images, sound, etc.).
- They have equalled or surpassed human performance over a wide variety of cognitively challenging tasks that had resisted earlier AI systems.
- They define the state of the art for most AI applications, and they have all but displaced symbolic systems.

The Deep Learning Revolution

- In the past three decades the emergence of powerful hardware (GPUs), the abundance of online data, and radical innovations in the architecture of neural networks, have produced the deep learning revolution.
- Transformers, which drive Large Language Models, consist entirely of blocks of attention heads.
- These are trained independently of each other, and they can identify fine grained patterns in data across distinct modalities (text, visual images, sound, etc.).
- They have equalled or surpassed human performance over a wide variety of cognitively challenging tasks that had resisted earlier AI systems.
- They define the state of the art for most AI applications, and they have all but displaced symbolic systems.

The Deep Learning Revolution

- In the past three decades the emergence of powerful hardware (GPUs), the abundance of online data, and radical innovations in the architecture of neural networks, have produced the deep learning revolution.
- Transformers, which drive Large Language Models, consist entirely of blocks of attention heads.
- These are trained independently of each other, and they can identify fine grained patterns in data across distinct modalities (text, visual images, sound, etc.).
- They have equalled or surpassed human performance over a wide variety of cognitively challenging tasks that had resisted earlier AI systems.
- They define the state of the art for most AI applications, and they have all but displaced symbolic systems.

The Deep Learning Revolution

- In the past three decades the emergence of powerful hardware (GPUs), the abundance of online data, and radical innovations in the architecture of neural networks, have produced the deep learning revolution.
- Transformers, which drive Large Language Models, consist entirely of blocks of attention heads.
- These are trained independently of each other, and they can identify fine grained patterns in data across distinct modalities (text, visual images, sound, etc.).
- They have equalled or surpassed human performance over a wide variety of cognitively challenging tasks that had resisted earlier AI systems.
- They define the state of the art for most AI applications, and they have all but displaced symbolic systems.

The Deep Learning Revolution

- In the past three decades the emergence of powerful hardware (GPUs), the abundance of online data, and radical innovations in the architecture of neural networks, have produced the deep learning revolution.
- Transformers, which drive Large Language Models, consist entirely of blocks of attention heads.
- These are trained independently of each other, and they can identify fine grained patterns in data across distinct modalities (text, visual images, sound, etc.).
- They have equalled or surpassed human performance over a wide variety of cognitively challenging tasks that had resisted earlier AI systems.
- They define the state of the art for most AI applications, and they have all but displaced symbolic systems.

Limitations of Large Language Models

- LLMs do not perform reliably on natural language inference (NLI) tasks, when subject to adversarial testing (Talman and Chatzikyriakidis, 2019; Talman et al., 2021).
- They also do not do well on many real world reasoning tasks (Mahowald et al., 2023).
- While transformers learn superficial patterns of inference and they are sensitive to some lexical semantic content in arguments, they do not acquire stable deep reasoning abilities.
- LLMs are notorious for hallucinating fluent but fictional content, which undermines their reliability for question-answering, and a variety of other applications.
- Transformers are computationally opaque, in large measure because their activation and probability generating functions (such as ReLU and softmax) are non-linear.

Limitations of Large Language Models

- LLMs do not perform reliably on natural language inference (NLI) tasks, when subject to adversarial testing (Talman and Chatzikyriakidis, 2019; Talman et al., 2021).
- They also do not do well on many real world reasoning tasks (Mahowald et al., 2023).
- While transformers learn superficial patterns of inference and they are sensitive to some lexical semantic content in arguments, they do not acquire stable deep reasoning abilities.
- LLMs are notorious for hallucinating fluent but fictional content, which undermines their reliability for question-answering, and a variety of other applications.
- Transformers are computationally opaque, in large measure because their activation and probability generating functions (such as ReLU and softmax) are non-linear.

Limitations of Large Language Models

- LLMs do not perform reliably on natural language inference (NLI) tasks, when subject to adversarial testing (Talman and Chatzikyriakidis, 2019; Talman et al., 2021).
- They also do not do well on many real world reasoning tasks (Mahowald et al., 2023).
- While transformers learn superficial patterns of inference and they are sensitive to some lexical semantic content in arguments, they do not acquire stable deep reasoning abilities.
- LLMs are notorious for hallucinating fluent but fictional content, which undermines their reliability for question-answering, and a variety of other applications.
- Transformers are computationally opaque, in large measure because their activation and probability generating functions (such as ReLU and softmax) are non-linear.

Limitations of Large Language Models

- LLMs do not perform reliably on natural language inference (NLI) tasks, when subject to adversarial testing (Talman and Chatzikyriakidis, 2019; Talman et al., 2021).
- They also do not do well on many real world reasoning tasks (Mahowald et al., 2023).
- While transformers learn superficial patterns of inference and they are sensitive to some lexical semantic content in arguments, they do not acquire stable deep reasoning abilities.
- LLMs are notorious for hallucinating fluent but fictional content, which undermines their reliability for question-answering, and a variety of other applications.
- Transformers are computationally opaque, in large measure because their activation and probability generating functions (such as ReLU and softmax) are non-linear.

Limitations of Large Language Models

- LLMs do not perform reliably on natural language inference (NLI) tasks, when subject to adversarial testing (Talman and Chatzikyriakidis, 2019; Talman et al., 2021).
- They also do not do well on many real world reasoning tasks (Mahowald et al., 2023).
- While transformers learn superficial patterns of inference and they are sensitive to some lexical semantic content in arguments, they do not acquire stable deep reasoning abilities.
- LLMs are notorious for hallucinating fluent but fictional content, which undermines their reliability for question-answering, and a variety of other applications.
- Transformers are computationally opaque, in large measure because their activation and probability generating functions (such as ReLU and softmax) are non-linear.

Advantages Claimed for Neuro-Symbolic Models

- Proponents of neuro-symbolic models assert that they significantly reduce training time by encoding information in symbolic features and rule systems, which would require additional data to extract.
- They argue that these models are more transparent than non-enriched DNNs, by virtue of the explainable nature of their symbolic content.
- They maintain that the symbolic component of these models substantially improves their performance, relative to non-symbolic DNNs, over a wide variety of tasks.
- In fact, the evidence for these claims is far from clear, in at least one major class of neuro-symbolic models.

Advantages Claimed for Neuro-Symbolic Models

- Proponents of neuro-symbolic models assert that they significantly reduce training time by encoding information in symbolic features and rule systems, which would require additional data to extract.
- They argue that these models are more transparent than non-enriched DNNs, by virtue of the explainable nature of their symbolic content.
- They maintain that the symbolic component of these models substantially improves their performance, relative to non-symbolic DNNs, over a wide variety of tasks.
- In fact, the evidence for these claims is far from clear, in at least one major class of neuro-symbolic models.

Advantages Claimed for Neuro-Symbolic Models

- Proponents of neuro-symbolic models assert that they significantly reduce training time by encoding information in symbolic features and rule systems, which would require additional data to extract.
- They argue that these models are more transparent than non-enriched DNNs, by virtue of the explainable nature of their symbolic content.
- They maintain that the symbolic component of these models substantially improves their performance, relative to non-symbolic DNNs, over a wide variety of tasks.
- In fact, the evidence for these claims is far from clear, in at least one major class of neuro-symbolic models.

Advantages Claimed for Neuro-Symbolic Models

- Proponents of neuro-symbolic models assert that they significantly reduce training time by encoding information in symbolic features and rule systems, which would require additional data to extract.
- They argue that these models are more transparent than non-enriched DNNs, by virtue of the explainable nature of their symbolic content.
- They maintain that the symbolic component of these models substantially improves their performance, relative to non-symbolic DNNs, over a wide variety of tasks.
- In fact, the evidence for these claims is far from clear, in at least one major class of neuro-symbolic models.

Injecting Symbolic Representations into a DNN

- Some theorists have revived Minksy's call for the development of hybrid neuro-symbolic models (Marcus, 2022).
- One way of constructing a hybrid framework is to inject symbolic representations into the processing operations of a Deep Neural Network (DNN).
- This can be done directly, by revising the architecture of the DNN to incorporate the biases of a symbolic system into its computation, at different levels of the network.
- Injection can also be achieved indirectly, through training the DNN on a biased distribution that a symbolic system generates (knowledge distillation).
- Symbolic markers, or structures, can also be inserted into the data on which a DNN is trained.

Injecting Symbolic Representations into a DNN

- Some theorists have revived Minksy's call for the development of hybrid neuro-symbolic models (Marcus, 2022).
- One way of constructing a hybrid framework is to inject symbolic representations into the processing operations of a Deep Neural Network (DNN).
- This can be done directly, by revising the architecture of the DNN to incorporate the biases of a symbolic system into its computation, at different levels of the network.
- Injection can also be achieved indirectly, through training the DNN on a biased distribution that a symbolic system generates (knowledge distillation).
- Symbolic markers, or structures, can also be inserted into the data on which a DNN is trained.

Injecting Symbolic Representations into a DNN

- Some theorists have revived Minksy's call for the development of hybrid neuro-symbolic models (Marcus, 2022).
- One way of constructing a hybrid framework is to inject symbolic representations into the processing operations of a Deep Neural Network (DNN).
- This can be done directly, by revising the architecture of the DNN to incorporate the biases of a symbolic system into its computation, at different levels of the network.
- Injection can also be achieved indirectly, through training the DNN on a biased distribution that a symbolic system generates (knowledge distillation).
- Symbolic markers, or structures, can also be inserted into the data on which a DNN is trained.

Injecting Symbolic Representations into a DNN

- Some theorists have revived Minksy's call for the development of hybrid neuro-symbolic models (Marcus, 2022).
- One way of constructing a hybrid framework is to inject symbolic representations into the processing operations of a Deep Neural Network (DNN).
- This can be done directly, by revising the architecture of the DNN to incorporate the biases of a symbolic system into its computation, at different levels of the network.
- Injection can also be achieved indirectly, through training the DNN on a biased distribution that a symbolic system generates (knowledge distillation).
- Symbolic markers, or structures, can also be inserted into the data on which a DNN is trained.

Injecting Symbolic Representations into a DNN

- Some theorists have revived Minksy's call for the development of hybrid neuro-symbolic models (Marcus, 2022).
- One way of constructing a hybrid framework is to inject symbolic representations into the processing operations of a Deep Neural Network (DNN).
- This can be done directly, by revising the architecture of the DNN to incorporate the biases of a symbolic system into its computation, at different levels of the network.
- Injection can also be achieved indirectly, through training the DNN on a biased distribution that a symbolic system generates (knowledge distillation).
- Symbolic markers, or structures, can also be inserted into the data on which a DNN is trained.

Tree DNNs for NLP

- Tree DNNs incorporate syntactic structure into a Deep Neural Network (DNN), either directly through its architecture, or indirectly through knowledge distillation and training data.
- Socher et al. (2011), Bowman et al. (2016), Yogatama et al. (2017), Choi et al. (2018), Williams et al. (2018), Maillard et al. (2019), Ek et al. (2019) consider LSTM-based Tree DNNs.
- These have been applied to NLP tasks like sentiment analysis, NLI, and the prediction of human sentence acceptability judgments.
- They have yielded small improvements in performance, which do not provide strong motivation for inserting trees, or syntactic and semantic markers into LSTMs.

Tree DNNs for NLP

- Tree DNNs incorporate syntactic structure into a Deep Neural Network (DNN), either directly through its architecture, or indirectly through knowledge distillation and training data.
- Socher et al. (2011), Bowman et al. (2016), Yogatama et al. (2017), Choi et al. (2018), Williams et al. (2018), Maillard et al. (2019), Ek et al. (2019) consider LSTM-based Tree DNNs.
- These have been applied to NLP tasks like sentiment analysis, NLI, and the prediction of human sentence acceptability judgments.
- They have yielded small improvements in performance, which do not provide strong motivation for inserting trees, or syntactic and semantic markers into LSTMs.

Tree DNNs for NLP

- Tree DNNs incorporate syntactic structure into a Deep Neural Network (DNN), either directly through its architecture, or indirectly through knowledge distillation and training data.
- Socher et al. (2011), Bowman et al. (2016), Yogatama et al. (2017), Choi et al. (2018), Williams et al. (2018), Maillard et al. (2019), Ek et al. (2019) consider LSTM-based Tree DNNs.
- These have been applied to NLP tasks like sentiment analysis, NLI, and the prediction of human sentence acceptability judgments.
- They have yielded small improvements in performance, which do not provide strong motivation for inserting trees, or syntactic and semantic markers into LSTMs.

Tree DNNs for NLP

- Tree DNNs incorporate syntactic structure into a Deep Neural Network (DNN), either directly through its architecture, or indirectly through knowledge distillation and training data.
- Socher et al. (2011), Bowman et al. (2016), Yogatama et al. (2017), Choi et al. (2018), Williams et al. (2018), Maillard et al. (2019), Ek et al. (2019) consider LSTM-based Tree DNNs.
- These have been applied to NLP tasks like sentiment analysis, NLI, and the prediction of human sentence acceptability judgments.
- They have yielded small improvements in performance, which do not provide strong motivation for inserting trees, or syntactic and semantic markers into LSTMs.

Syntactic Enrichment of BERT

- More recent work has incorporated syntactic tree structure into transformers like BERT, and applied them to a broader range of tasks.
- Bai et al. (2021) integrate tree structure recognition into the attention head blocks of BERT and RoBERTa.
- They test different versions of these transformers on the GLUE benchmark tasks, which include sentence acceptability assessment, paraphrase recognition, and NLI.
- For the overwhelming majority of cases they report an accuracy gain of the tree enriched model, relative to its non-tree counterpart, of between 1% and 2%.
- These results suggest that the contribution of the implemented tree structure enrichment to BERT and RoBERTa 's performance on the GLUE tasks is marginal.

Syntactic Enrichment of BERT

- More recent work has incorporated syntactic tree structure into transformers like BERT, and applied them to a broader range of tasks.
- Bai et al. (2021) integrate tree structure recognition into the attention head blocks of BERT and RoBERTa.
- They test different versions of these transformers on the GLUE benchmark tasks, which include sentence acceptability assessment, paraphrase recognition, and NLI.
- For the overwhelming majority of cases they report an accuracy gain of the tree enriched model, relative to its non-tree counterpart, of between 1% and 2%.
- These results suggest that the contribution of the implemented tree structure enrichment to BERT and RoBERTa 's performance on the GLUE tasks is marginal.

Syntactic Enrichment of BERT

- More recent work has incorporated syntactic tree structure into transformers like BERT, and applied them to a broader range of tasks.
- Bai et al. (2021) integrate tree structure recognition into the attention head blocks of BERT and RoBERTa.
- They test different versions of these transformers on the GLUE benchmark tasks, which include sentence acceptability assessment, paraphrase recognition, and NLI.
- For the overwhelming majority of cases they report an accuracy gain of the tree enriched model, relative to its non-tree counterpart, of between 1% and 2%.
- These results suggest that the contribution of the implemented tree structure enrichment to BERT and RoBERTa 's performance on the GLUE tasks is marginal.

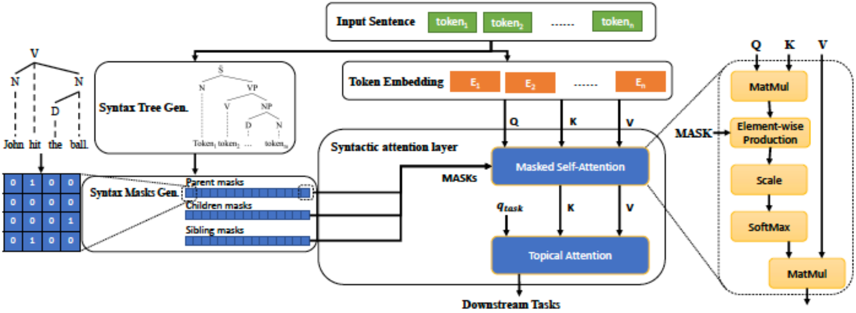
Syntactic Enrichment of BERT

- More recent work has incorporated syntactic tree structure into transformers like BERT, and applied them to a broader range of tasks.
- Bai et al. (2021) integrate tree structure recognition into the attention head blocks of BERT and RoBERTa.
- They test different versions of these transformers on the GLUE benchmark tasks, which include sentence acceptability assessment, paraphrase recognition, and NLI.
- For the overwhelming majority of cases they report an accuracy gain of the tree enriched model, relative to its non-tree counterpart, of between 1% and 2%.
- These results suggest that the contribution of the implemented tree structure enrichment to BERT and RoBERTa 's performance on the GLUE tasks is marginal.

Syntactic Enrichment of BERT

- More recent work has incorporated syntactic tree structure into transformers like BERT, and applied them to a broader range of tasks.
- Bai et al. (2021) integrate tree structure recognition into the attention head blocks of BERT and RoBERTa.
- They test different versions of these transformers on the GLUE benchmark tasks, which include sentence acceptability assessment, paraphrase recognition, and NLI.
- For the overwhelming majority of cases they report an accuracy gain of the tree enriched model, relative to its non-tree counterpart, of between 1% and 2%.
- These results suggest that the contribution of the implemented tree structure enrichment to BERT and RoBERTa ’s performance on the GLUE tasks is marginal.

Syntax-BERT: Bai et al. (2021)



0	1	0	0
0	0	0	0
0	0	0	1
0	1	0	0

Adding Dependency Tree Graphs to BERT

- Sachan et al. (2021) enrich BERT and RoBERTa with dependency tree graphs.
- They test them on semantic role labelling, named entity recognition, and relation extraction.
- For in domain test sets the graph versions of the models achieve F1 scores that are 1%-2% higher than their non-enriched counterparts.
- In an out of domain test on semantic role labelling, the gain in F1 score was 2%-5%.
- These results are similar to those that Bai et al. (2021) report for their syntactic tree versions of BERT and RoBERTa.

Adding Dependency Tree Graphs to BERT

- Sachan et al. (2021) enrich BERT and RoBERTa with dependency tree graphs.
- They test them on semantic role labelling, named entity recognition, and relation extraction.
- For in domain test sets the graph versions of the models achieve F1 scores that are 1%-2% higher than their non-enriched counterparts.
- In an out of domain test on semantic role labelling, the gain in F1 score was 2%-5%.
- These results are similar to those that Bai et al. (2021) report for their syntactic tree versions of BERT and RoBERTa.

Adding Dependency Tree Graphs to BERT

- Sachan et al. (2021) enrich BERT and RoBERTa with dependency tree graphs.
- They test them on semantic role labelling, named entity recognition, and relation extraction.
- For in domain test sets the graph versions of the models achieve F1 scores that are 1%-2% higher than their non-enriched counterparts.
- In an out of domain test on semantic role labelling, the gain in F1 score was 2%-5%.
- These results are similar to those that Bai et al. (2021) report for their syntactic tree versions of BERT and RoBERTa.

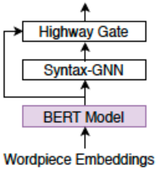
Adding Dependency Tree Graphs to BERT

- Sachan et al. (2021) enrich BERT and RoBERTa with dependency tree graphs.
- They test them on semantic role labelling, named entity recognition, and relation extraction.
- For in domain test sets the graph versions of the models achieve F1 scores that are 1%-2% higher than their non-enriched counterparts.
- In an out of domain test on semantic role labelling, the gain in F1 score was 2%-5%.
- These results are similar to those that Bai et al. (2021) report for their syntactic tree versions of BERT and RoBERTa.

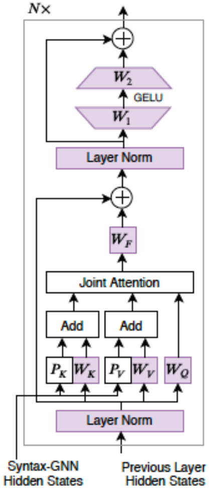
Adding Dependency Tree Graphs to BERT

- Sachan et al. (2021) enrich BERT and RoBERTa with dependency tree graphs.
- They test them on semantic role labelling, named entity recognition, and relation extraction.
- For in domain test sets the graph versions of the models achieve F1 scores that are 1%-2% higher than their non-enriched counterparts.
- In an out of domain test on semantic role labelling, the gain in F1 score was 2%-5%.
- These results are similar to those that Bai et al. (2021) report for their syntactic tree versions of BERT and RoBERTa.

Dependency Tree Graphs Bert: Sachan et al. (2021)



(a) Late Fusion



(b) Joint Fusion

An Injective Model for Medical Image Identification

- Abdullah et al. (2023) enrich a CNN by infusing handcrafted knowledge features for segmenting brain aneurism images.
- They experiment with feature infusion at different levels of the network.
- They use Intersection over Union (IoU) as the metric to compare several versions of the feature infused CNN with its non-enriched baseline.
- $\text{IoU} = \frac{area(M_{image}) \cap area(GT_{image})}{area(M_{image}) \cup area(GT_{image})}$
- Their best feature infused model scored an IoU of 0.9676, while the non-enriched CNN achieved 0.9158.

An Injective Model for Medical Image Identification

- Abdullah et al. (2023) enrich a CNN by infusing handcrafted knowledge features for segmenting brain aneurism images.
- They experiment with feature infusion at different levels of the network.
- They use Intersection over Union (IoU) as the metric to compare several versions of the feature infused CNN with its non-enriched baseline.
- IoU = $\frac{area(M_{image}) \cap area(GT_{image})}{area(M_{image}) \cup area(GT_{image})}$
- Their best feature infused model scored an IoU of 0.9676, while the non-enriched CNN achieved 0.9158.

An Injective Model for Medical Image Identification

- Abdullah et al. (2023) enrich a CNN by infusing handcrafted knowledge features for segmenting brain aneurism images.
- They experiment with feature infusion at different levels of the network.
- They use Intersection over Union (IoU) as the metric to compare several versions of the feature infused CNN with its non-enriched baseline.
- $\mathrm{IoU} = \frac{area(M_{image}) \cap area(GT_{image})}{area(M_{image}) \cup area(GT_{image})}$
- Their best feature infused model scored an IoU of 0.9676, while the non-enriched CNN achieved 0.9158.

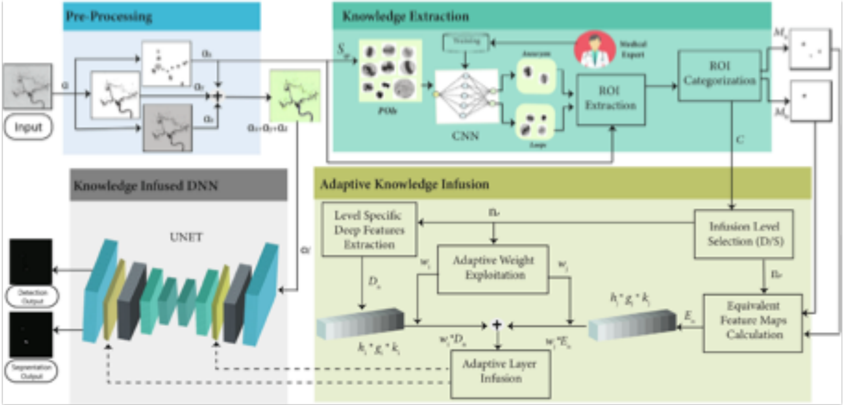
An Injective Model for Medical Image Identification

- Abdullah et al. (2023) enrich a CNN by infusing handcrafted knowledge features for segmenting brain aneurism images.
- They experiment with feature infusion at different levels of the network.
- They use Intersection over Union (IoU) as the metric to compare several versions of the feature infused CNN with its non-enriched baseline.
- IoU $= \dfrac{area(M_{image}) \cap area(GT_{image})}{area(M_{image}) \cup area(GT_{image})}$
- Their best feature infused model scored an IoU of 0.9676, while the non-enriched CNN achieved 0.9158.

An Injective Model for Medical Image Identification

- Abdullah et al. (2023) enrich a CNN by infusing handcrafted knowledge features for segmenting brain aneurism images.
- They experiment with feature infusion at different levels of the network.
- They use Intersection over Union (IoU) as the metric to compare several versions of the feature infused CNN with its non-enriched baseline.
- IoU = $\frac{area(M_{image}) \cap area(GT_{image})}{area(M_{image}) \cup area(GT_{image})}$
- Their best feature infused model scored an IoU of 0.9676, while the non-enriched CNN achieved 0.9158.

Knowledge Feature Infused CNN: Abdullah et al. (2023)



An Injective Model for Diabetes Diagnosis

- Lu et al. (2025) modify the hidden units of a CNN to function as probabilistic logical operators.
- They train the network to extract rules for diagnosing diabetes on the basis of data encoded as feature vectors.
- They compare alternative implementations of their rule learning CNN with traditional machine learning methods used for medical diagnosis.
- Their highest scoring model obtains an F1 score of 0.6875 on their test set, while they report Random Forest as achieving the best traditional ML result at 0.6380.
- Their best enriched CNN for AUC binary classification scores 0.8457, while Random Forest achieves 0.8342.
- Interestingly, they do not provide a comparison between their logically enriched CNN and a baseline version of the same model.

An Injective Model for Diabetes Diagnosis

- Lu et al. (2025) modify the hidden units of a CNN to function as probabilistic logical operators.
- They train the network to extract rules for diagnosing diabetes on the basis of data encoded as feature vectors.
- They compare alternative implementations of their rule learning CNN with traditional machine learning methods used for medical diagnosis.
- Their highest scoring model obtains an F1 score of 0.6875 on their test set, while they report Random Forest as achieving the best traditional ML result at 0.6380.
- Their best enriched CNN for AUC binary classification scores 0.8457, while Random Forest achieves 0.8342.
- Interestingly, they do not provide a comparison between their logically enriched CNN and a baseline version of the same model.

An Injective Model for Diabetes Diagnosis

- Lu et al. (2025) modify the hidden units of a CNN to function as probabilistic logical operators.
- They train the network to extract rules for diagnosing diabetes on the basis of data encoded as feature vectors.
- They compare alternative implementations of their rule learning CNN with traditional machine learning methods used for medical diagnosis.
- Their highest scoring model obtains an F1 score of 0.6875 on their test set, while they report Random Forest as achieving the best traditional ML result at 0.6380.
- Their best enriched CNN for AUC binary classification scores 0.8457, while Random Forest achieves 0.8342.
- Interestingly, they do not provide a comparison between their logically enriched CNN and a baseline version of the same model.

An Injective Model for Diabetes Diagnosis

- Lu et al. (2025) modify the hidden units of a CNN to function as probabilistic logical operators.
- They train the network to extract rules for diagnosing diabetes on the basis of data encoded as feature vectors.
- They compare alternative implementations of their rule learning CNN with traditional machine learning methods used for medical diagnosis.
- Their highest scoring model obtains an F1 score of 0.6875 on their test set, while they report Random Forest as achieving the best traditional ML result at 0.6380.
- Their best enriched CNN for AUC binary classification scores 0.8457, while Random Forest achieves 0.8342.
- Interestingly, they do not provide a comparison between their logically enriched CNN and a baseline version of the same model.

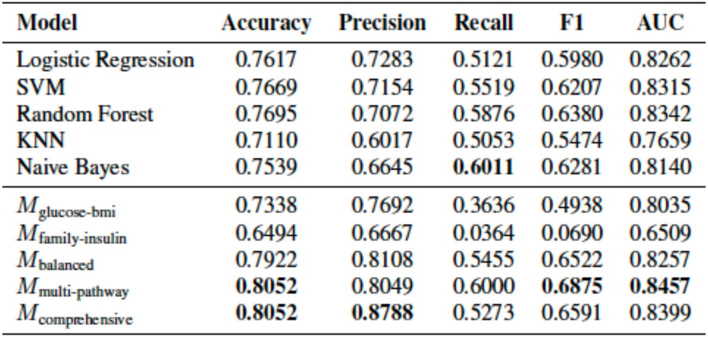
An Injective Model for Diabetes Diagnosis

- Lu et al. (2025) modify the hidden units of a CNN to function as probabilistic logical operators.
- They train the network to extract rules for diagnosing diabetes on the basis of data encoded as feature vectors.
- They compare alternative implementations of their rule learning CNN with traditional machine learning methods used for medical diagnosis.
- Their highest scoring model obtains an F1 score of 0.6875 on their test set, while they report Random Forest as achieving the best traditional ML result at 0.6380.
- Their best enriched CNN for AUC binary classification scores 0.8457, while Random Forest achieves 0.8342.
- Interestingly, they do not provide a comparison between their logically enriched CNN and a baseline version of the same model.

An Injective Model for Diabetes Diagnosis

- Lu et al. (2025) modify the hidden units of a CNN to function as probabilistic logical operators.
- They train the network to extract rules for diagnosing diabetes on the basis of data encoded as feature vectors.
- They compare alternative implementations of their rule learning CNN with traditional machine learning methods used for medical diagnosis.
- Their highest scoring model obtains an F1 score of 0.6875 on their test set, while they report Random Forest as achieving the best traditional ML result at 0.6380.
- Their best enriched CNN for AUC binary classification scores 0.8457, while Random Forest achieves 0.8342.
- Interestingly, they do not provide a comparison between their logically enriched CNN and a baseline version of the same model.

Lu et al. (2023) Experimental Results



Model	**Accuracy**	**Precision**	**Recall**	**F1**	**AUC**
Logistic Regression	0.7617	0.7283	0.5121	0.5980	0.8262
SVM	0.7669	0.7154	0.5519	0.6207	0.8315
Random Forest	0.7695	0.7072	0.5876	0.6380	0.8342
KNN	0.7110	0.6017	0.5053	0.5474	0.7659
Naive Bayes	0.7539	0.6645	**0.6011**	0.6281	0.8140
$M_{\text{glucose-bmi}}$	0.7338	0.7692	0.3636	0.4938	0.8035
$M_{\text{family-insulin}}$	0.6494	0.6667	0.0364	0.0690	0.6509
M_{balanced}	0.7922	0.8108	0.5455	0.6522	0.8257
$M_{\text{multi-pathway}}$	**0.8052**	0.8049	0.6000	**0.6875**	**0.8457**
$M_{\text{comprehensive}}$	**0.8052**	**0.8788**	0.5273	0.6591	0.8399

Limitations of Injective Models

- Injective models provide small gains in performance relative to their unenriched counterparts.
- These gains tend to diminish with additional training data for non-enriched DNNs.
- The claim that injective models offer greater transparency than non-injective DNNs is open to question.
- In most cases injective DNNs remain non-compositional in their output at each level, as they continue to use non-linear functions like ReLU and softmax to generate output vectors.

Limitations of Injective Models

- Injective models provide small gains in performance relative to their unenriched counterparts.
- These gains tend to diminish with additional training data for non-enriched DNNs.
- The claim that injective models offer greater transparency than non-injective DNNs is open to question.
- In most cases injective DNNs remain non-compositional in their output at each level, as they continue to use non-linear functions like ReLU and softmax to generate output vectors.

Limitations of Injective Models

- Injective models provide small gains in performance relative to their unenriched counterparts.
- These gains tend to diminish with additional training data for non-enriched DNNs.
- The claim that injective models offer greater transparency than non-injective DNNs is open to question.
- In most cases injective DNNs remain non-compositional in their output at each level, as they continue to use non-linear functions like ReLU and softmax to generate output vectors.

Limitations of Injective Models

- Injective models provide small gains in performance relative to their unenriched counterparts.
- These gains tend to diminish with additional training data for non-enriched DNNs.
- The claim that injective models offer greater transparency than non-injective DNNs is open to question.
- In most cases injective DNNs remain non-compositional in their output at each level, as they continue to use non-linear functions like ReLU and softmax to generate output vectors.

Possible Reasons for the Limited Success of Injective Models

- Advocates of injective models tend to assume that humans acquire and represent most knowledge as rule sets that are best modelled as algebraic systems (grammars, logics, etc.).
- It is far from obvious that this is the case for all types of knowledge.
- It is entirely possible that humans encode many aspects of their discriminatory classification knowledge in non-symbolic, distributed representations of regularities (Smolensky, 1987; McClelland, 2016, among others).
- It is also possible that, by virtue of their design, DNNs are unable to easily integrate symbolic components into their distributed representations of information, in a way that significantly improves learning or inference.

Possible Reasons for the Limited Success of Injective Models

- Advocates of injective models tend to assume that humans acquire and represent most knowledge as rule sets that are best modelled as algebraic systems (grammars, logics, etc.).
- It is far from obvious that this is the case for all types of knowledge.
- It is entirely possible that humans encode many aspects of their discriminatory classification knowledge in non-symbolic, distributed representations of regularities (Smolensky, 1987; McClelland, 2016, among others).
- It is also possible that, by virtue of their design, DNNs are unable to easily integrate symbolic components into their distributed representations of information, in a way that significantly improves learning or inference.

Possible Reasons for the Limited Success of Injective Models

- Advocates of injective models tend to assume that humans acquire and represent most knowledge as rule sets that are best modelled as algebraic systems (grammars, logics, etc.).
- It is far from obvious that this is the case for all types of knowledge.
- It is entirely possible that humans encode many aspects of their discriminatory classification knowledge in non-symbolic, distributed representations of regularities (Smolensky, 1987; McClelland, 2016, among others).
- It is also possible that, by virtue of their design, DNNs are unable to easily integrate symbolic components into their distributed representations of information, in a way that significantly improves learning or inference.

Possible Reasons for the Limited Success of Injective Models

- Advocates of injective models tend to assume that humans acquire and represent most knowledge as rule sets that are best modelled as algebraic systems (grammars, logics, etc.).
- It is far from obvious that this is the case for all types of knowledge.
- It is entirely possible that humans encode many aspects of their discriminatory classification knowledge in non-symbolic, distributed representations of regularities (Smolensky, 1987; McClelland, 2016, among others).
- It is also possible that, by virtue of their design, DNNs are unable to easily integrate symbolic components into their distributed representations of information, in a way that significantly improves learning or inference.

A Federative Alternative to Injection Models

- A federative hybrid model does not inject symbolic content into a DNN.
- It combines a DNN with a symbolic reasoning module within a framework in which each of these systems functions autonomously.
- The framework sustains the distinct computational procedures that its two central components apply for representing information.
- In one version of this architecture the DNN extracts features for an interface that labels them, and feeds them to a logic based inference program.
- This approach seems closer than an injection model to Minsky's original proposal.

A Federative Alternative to Injection Models

- A federative hybrid model does not inject symbolic content into a DNN.
- It combines a DNN with a symbolic reasoning module within a framework in which each of these systems functions autonomously.
- The framework sustains the distinct computational procedures that its two central components apply for representing information.
- In one version of this architecture the DNN extracts features for an interface that labels them, and feeds them to a logic based inference program.
- This approach seems closer than an injection model to Minsky's original proposal.

A Federative Alternative to Injection Models

- A federative hybrid model does not inject symbolic content into a DNN.
- It combines a DNN with a symbolic reasoning module within a framework in which each of these systems functions autonomously.
- The framework sustains the distinct computational procedures that its two central components apply for representing information.
- In one version of this architecture the DNN extracts features for an interface that labels them, and feeds them to a logic based inference program.
- This approach seems closer than an injection model to Minsky's original proposal.

A Federative Alternative to Injection Models

- A federative hybrid model does not inject symbolic content into a DNN.
- It combines a DNN with a symbolic reasoning module within a framework in which each of these systems functions autonomously.
- The framework sustains the distinct computational procedures that its two central components apply for representing information.
- In one version of this architecture the DNN extracts features for an interface that labels them, and feeds them to a logic based inference program.
- This approach seems closer than an injection model to Minsky's original proposal.

A Federative Alternative to Injection Models

- A federative hybrid model does not inject symbolic content into a DNN.
- It combines a DNN with a symbolic reasoning module within a framework in which each of these systems functions autonomously.
- The framework sustains the distinct computational procedures that its two central components apply for representing information.
- In one version of this architecture the DNN extracts features for an interface that labels them, and feeds them to a logic based inference program.
- This approach seems closer than an injection model to Minsky's original proposal.

A Federative Model for Complex Image Identification

- Cunnington et al. (2023) present a Feed Forward Neural-Symbolic Learner (FFNSL) for image classification.
- It consists of a DNN for extracting features from images, an interface component that assigns labels to these features, and a logic based system (an ILP) that learns rules from these labelled features.
- They test variants of this model on a suite of image classification tasks in which knowledge of a game, or a problem, are necessary for the correct solution.
- They use distributional shifts of training and test data (through image rotation) to ascertain the robustness of the system under variation.

A Federative Model for Complex Image Identification

- Cunnington et al. (2023) present a Feed Forward Neural-Symbolic Learner (FFNSL) for image classification.
- It consists of a DNN for extracting features from images, an interface component that assigns labels to these features, and a logic based system (an ILP) that learns rules from these labelled features.
- They test variants of this model on a suite of image classification tasks in which knowledge of a game, or a problem, are necessary for the correct solution.
- They use distributional shifts of training and test data (through image rotation) to ascertain the robustness of the system under variation.

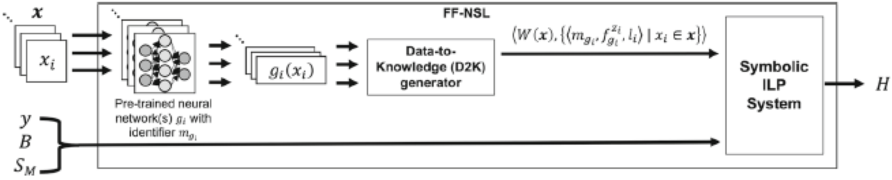
A Federative Model for Complex Image Identification

- Cunnington et al. (2023) present a Feed Forward Neural-Symbolic Learner (FFNSL) for image classification.
- It consists of a DNN for extracting features from images, an interface component that assigns labels to these features, and a logic based system (an ILP) that learns rules from these labelled features.
- They test variants of this model on a suite of image classification tasks in which knowledge of a game, or a problem, are necessary for the correct solution.
- They use distributional shifts of training and test data (through image rotation) to ascertain the robustness of the system under variation.

A Federative Model for Complex Image Identification

- Cunnington et al. (2023) present a Feed Forward Neural-Symbolic Learner (FFNSL) for image classification.
- It consists of a DNN for extracting features from images, an interface component that assigns labels to these features, and a logic based system (an ILP) that learns rules from these labelled features.
- They test variants of this model on a suite of image classification tasks in which knowledge of a game, or a problem, are necessary for the correct solution.
- They use distributional shifts of training and test data (through image rotation) to ascertain the robustness of the system under variation.

Architecture of FFNSL


$\boldsymbol{x}$
x_i
FF-NSL
Pre-trained neural network(s) g_i with identifier m_{g_i}
$g_i(x_i)$
Data-to-Knowledge (D2K) generator
$\langle W(\boldsymbol{x}), \{\langle m_{g_i}, f_{g_i}^{z_i}, l_i\rangle \mid x_i \in \boldsymbol{x}\}\rangle$
Symbolic ILP System
H
y
B
S_M

Performance of FFNLS

- FFNSL models exhibit significant gains over non-symbolic ML and DNN baselines.
- They require significantly less training data to achieve high accuracy in complex image classification tasks.
- They remain stable over higher levels of distributional shift in both training and test data.
- They generate transparent rule-based hypotheses.

Performance of FFNLS

- FFNSL models exhibit significant gains over non-symbolic ML and DNN baselines.
- They require significantly less training data to achieve high accuracy in complex image classification tasks.
- They remain stable over higher levels of distributional shift in both training and test data.
- They generate transparent rule-based hypotheses.

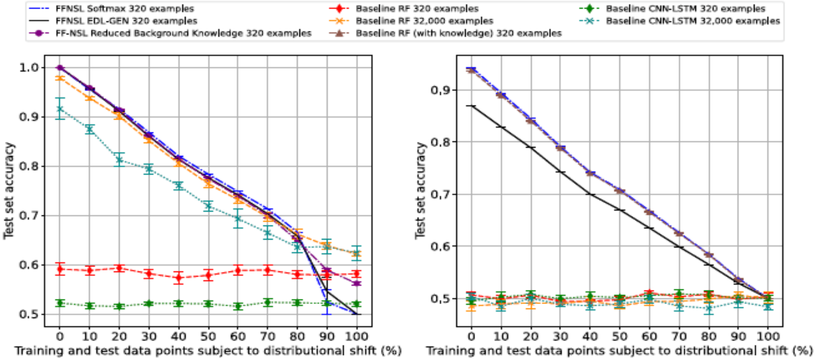
Performance of FFNLS

- FFNSL models exhibit significant gains over non-symbolic ML and DNN baselines.
- They require significantly less training data to achieve high accuracy in complex image classification tasks.
- They remain stable over higher levels of distributional shift in both training and test data.
- They generate transparent rule-based hypotheses.

Performance of FFNLS

- FFNSL models exhibit significant gains over non-symbolic ML and DNN baselines.
- They require significantly less training data to achieve high accuracy in complex image classification tasks.
- They remain stable over higher levels of distributional shift in both training and test data.
- They generate transparent rule-based hypotheses.

FFNSL Accuracy on Sudoku Grid Validity Recognition



(a) 4 × 4 grids (b) 9 × 9 grids

Conclusions

- The injection of symbolic features or rule-based biases directly into a DNN does not seem to signficantly improve its performance, relative to a non-enriched version of the same model.
- This may be due to the differences in the way DNNs and symbolic systems represent patterns of regularity.
- Federative neuro-symbolic models sustain the internal integrity and autonomy of both types of processing system.
- They appear to offer a more effective way of combining the strengths of each framework.
- Further research on both injective and federative models is required to ascertain the extent to which this conjecture holds.

Conclusions

- The injection of symbolic features or rule-based biases directly into a DNN does not seem to signficantly improve its performance, relative to a non-enriched version of the same model.
- This may be due to the differences in the way DNNs and symbolic systems represent patterns of regularity.
- Federative neuro-symbolic models sustain the internal integrity and autonomy of both types of processing system.
- They appear to offer a more effective way of combining the strengths of each framework.
- Further research on both injective and federative models is required to ascertain the extent to which this conjecture holds.

Conclusions

- The injection of symbolic features or rule-based biases directly into a DNN does not seem to signficantly improve its performance, relative to a non-enriched version of the same model.
- This may be due to the differences in the way DNNs and symbolic systems represent patterns of regularity.
- Federative neuro-symbolic models sustain the internal integrity and autonomy of both types of processing system.
- They appear to offer a more effective way of combining the strengths of each framework.
- Further research on both injective and federative models is required to ascertain the extent to which this conjecture holds.

Conclusions

- The injection of symbolic features or rule-based biases directly into a DNN does not seem to signficantly improve its performance, relative to a non-enriched version of the same model.
- This may be due to the differences in the way DNNs and symbolic systems represent patterns of regularity.
- Federative neuro-symbolic models sustain the internal integrity and autonomy of both types of processing system.
- They appear to offer a more effective way of combining the strengths of each framework.
- Further research on both injective and federative models is required to ascertain the extent to which this conjecture holds.

Conclusions

- The injection of symbolic features or rule-based biases directly into a DNN does not seem to signficantly improve its performance, relative to a non-enriched version of the same model.
- This may be due to the differences in the way DNNs and symbolic systems represent patterns of regularity.
- Federative neuro-symbolic models sustain the internal integrity and autonomy of both types of processing system.
- They appear to offer a more effective way of combining the strengths of each framework.
- Further research on both injective and federative models is required to ascertain the extent to which this conjecture holds.

Future Work

- More extensive comparisons of injective and non-injective state of the art transformers, over a wider variety of tasks, is needed to obtain a better sense of the limits of this approach.
- Similarly, federative models in which current transformers are used as the the DNN, with testing against the unenriched transformers, will help to clarify the prospects of this version of neuro-symbolic machine learning.
- At this point, federative models may present the most efficient way of augmenting the reasoning and inference capacities of DNNs.
- They also suggest a route to greater transparency in DNN driven machine learning.

Future Work

- More extensive comparisons of injective and non-injective state of the art transformers, over a wider variety of tasks, is needed to obtain a better sense of the limits of this approach.
- Similarly, federative models in which current transformers are used as the the DNN, with testing against the unenriched transformers, will help to clarify the prospects of this version of neuro-symbolic machine learning.
- At this point, federative models may present the most efficient way of augmenting the reasoning and inference capacities of DNNs.
- They also suggest a route to greater transparency in DNN driven machine learning.

Future Work

- More extensive comparisons of injective and non-injective state of the art transformers, over a wider variety of tasks, is needed to obtain a better sense of the limits of this approach.
- Similarly, federative models in which current transformers are used as the the DNN, with testing against the unenriched transformers, will help to clarify the prospects of this version of neuro-symbolic machine learning.
- At this point, federative models may present the most efficient way of augmenting the reasoning and inference capacities of DNNs.
- They also suggest a route to greater transparency in DNN driven machine learning.

Future Work

- More extensive comparisons of injective and non-injective state of the art transformers, over a wider variety of tasks, is needed to obtain a better sense of the limits of this approach.
- Similarly, federative models in which current transformers are used as the the DNN, with testing against the unenriched transformers, will help to clarify the prospects of this version of neuro-symbolic machine learning.
- At this point, federative models may present the most efficient way of augmenting the reasoning and inference capacities of DNNs.
- They also suggest a route to greater transparency in DNN driven machine learning.