Introduction to Multiple Instance Learning

Marc-André Carbonneau

Supervisors : Eric Granger and Ghyslain Gagnon

October 19th 2016

Outline of the presentation

1. Definition and formulation.
2. Applications.
3. Type of approaches
4. Characteristics of MIL problems.

What Is Multiple Instance Learning?

Problem Formulation

Multiple Instance Learning

What it is:

- It is a form of **weakly** supervised learning.
- Training instances are arranged in **sets**, called bags.
- A label is provided for entire bags **but not for instances**.

What it is not:

- Supervised learning
- Unsupervised learning
- Semi-supervised learning

Illustration of a MIL problem

Can enter the secret room

Can not enter the secret room

Can I the secret room???

What is the magic key???

Why use Multiple Instance Learning?

It has been proposed because:

- Some problems are **naturally formulated** as MIL

It is gaining momentum in the pattern recognition community because:

- It deals with **weakly annotated data**.
  - This reduces the annotation cost.
  - Algorithms can now learn from a greater quantity of training data.

Definition of the standard MIL assumption

- Training instances are arranged in sets generally called bags.
- A label is given to bags but not to individual instances.
- Negative bags do not contain positive instances.
- Positive bags may contain negative and positive instances.
- Positive bags contain at least one positive instance.

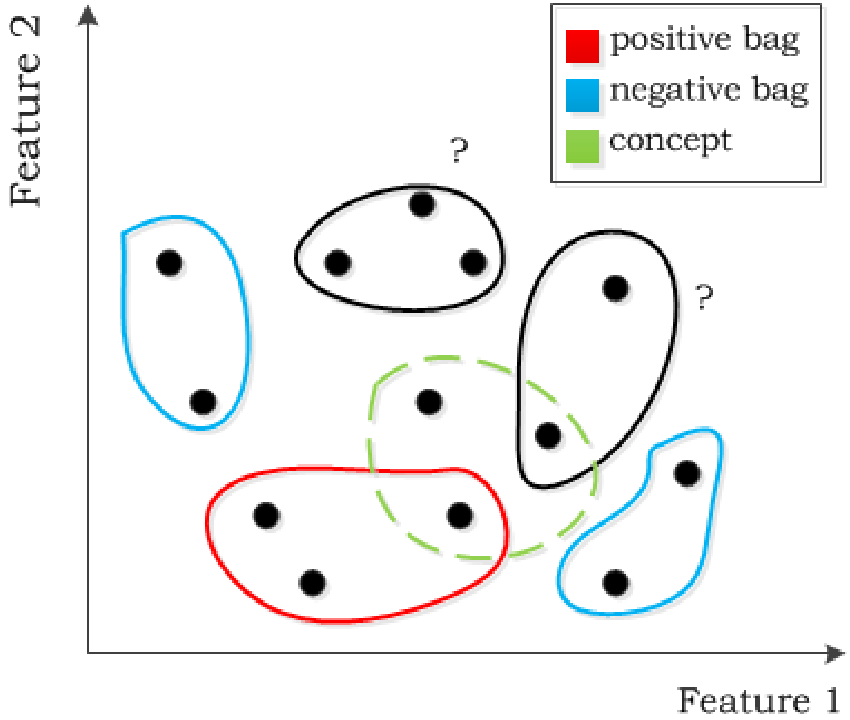


Image from : http://www.miproblems.org/mi-learning/

Relaxed MIL assumptions

In many applications, the standard MIL assumption is to restrictive. MIL can alternatively formulated as:

- A bag is positive when it contains a sufficient number of positive instances.
- A bag is positive when it contains a certain combination of positive instances.
- Positive and negative bags differ by their instance distributions.

More on MIL assumption: J. Foulds and E. Frank, “A Review of Multi-Instance Learning Assumptions,” *Knowl. Eng. Rev.*, vol. 25, no. 1, pp. 1–25, Mar. 2010.

Example of relaxed MIL assumptions

- Both **sand** and **water** segments are positive instances for beach pictures.
- However, picture of beach must contain **both segments** of sand and water. Otherwise, they can be pictures of desert or sea.

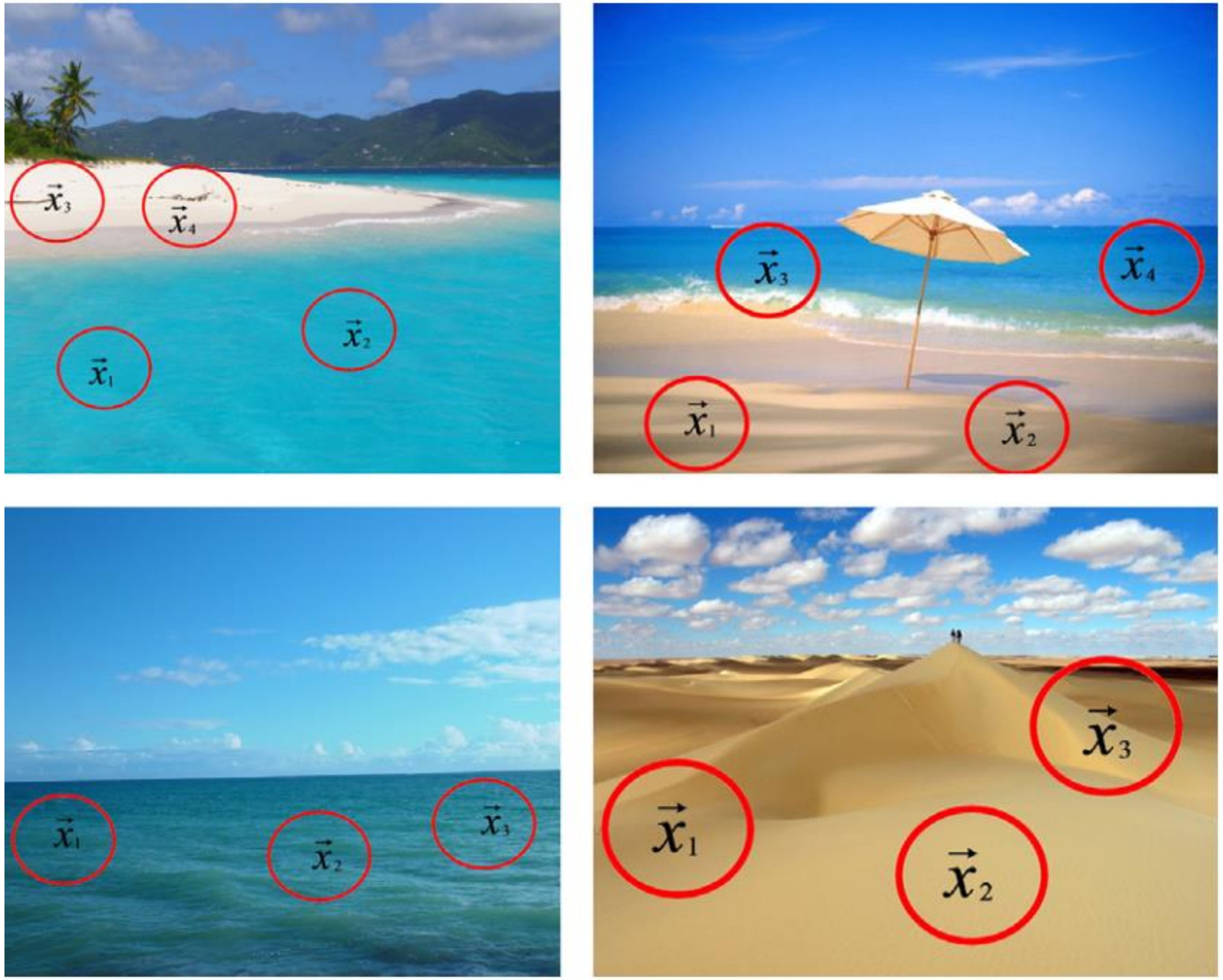


Image from : J. Amores, “Multiple instance classification: Review, taxonomy and comparative study,” *Artif. Intell.*, vol. 201, pp. 81–105, Aug. 2013.

Tasks that can be performed in MIL

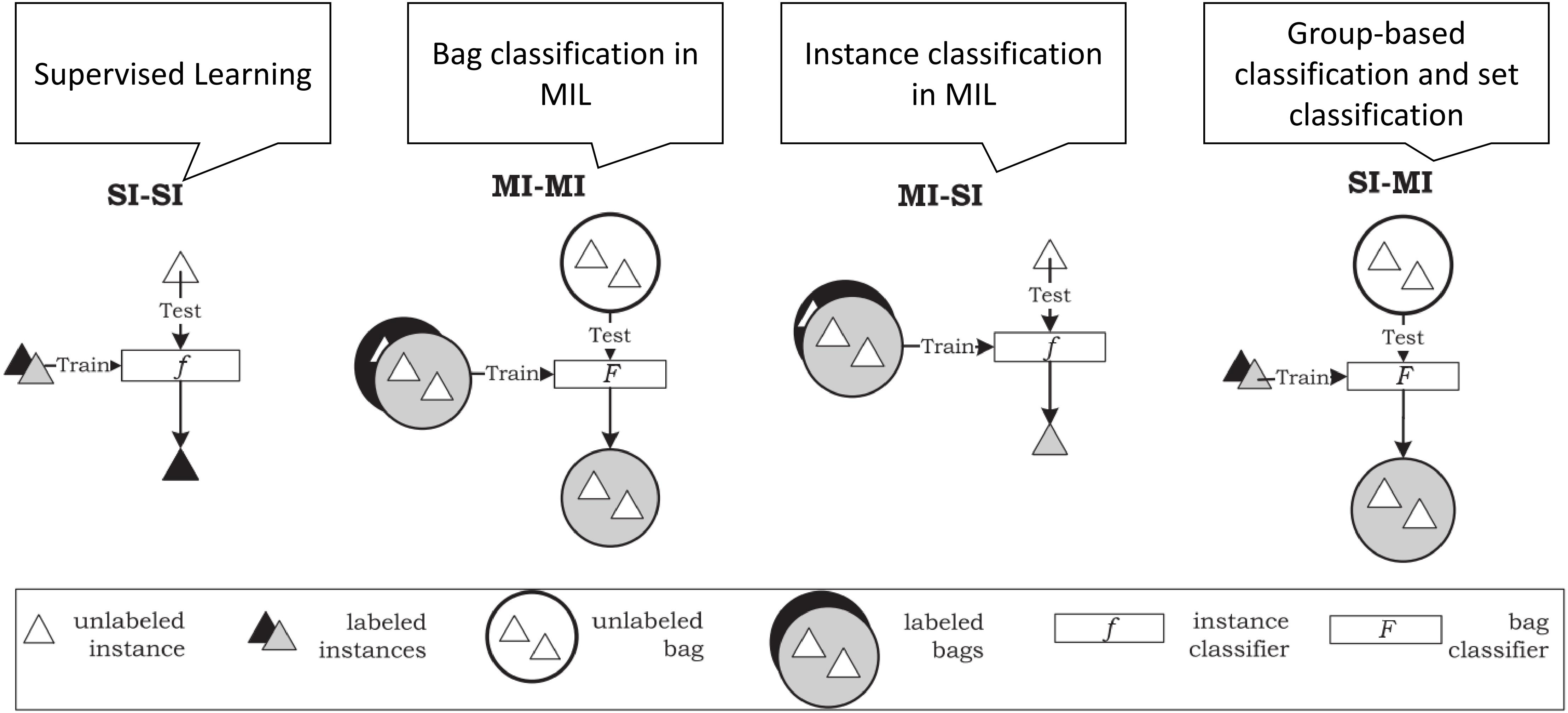
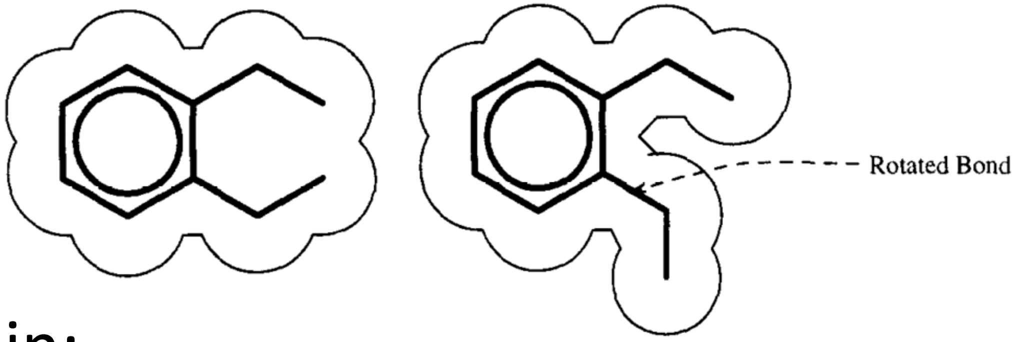


Image from: V. Cheplygina, D. M. J. Tax, and M. Loog, “On classification with bags, groups and sets,” *Pattern Recognition Letters*, vol. 59, pp. 11–17, Jul. 2015.

What Can I Do with Multiple Instance Learning?

Applications

Molecule Classification



This is the *first* MIL application published in:

T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the Multiple Instance Problem with Axis-parallel Rectangles," *Artificial Intelligence* 1997.

Objective: Predict if a molecule produces a given effect.

Bag: Collection of all conformations of the same molecule.

Instance: Conformation of a molecule.

Justification: Conformations are not observable individually.

Content Base Image Retrieval

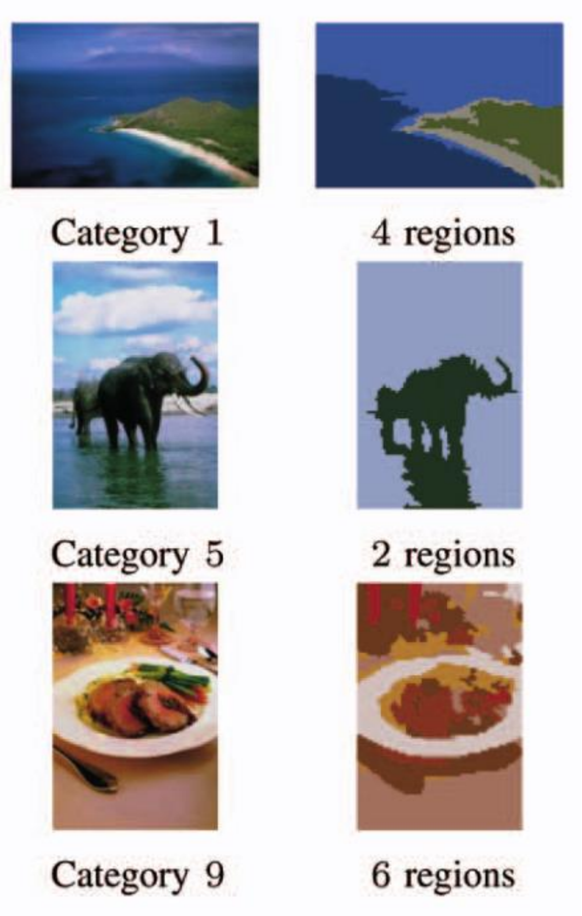
Objective: Classify images based on their subject.

Bag: Collection segments or patches extracted from an image.

Instance: Image segments or patches.

Justification: Images can represent composite objects or concepts.

Note: Bag-of-words methods are MIL methods.



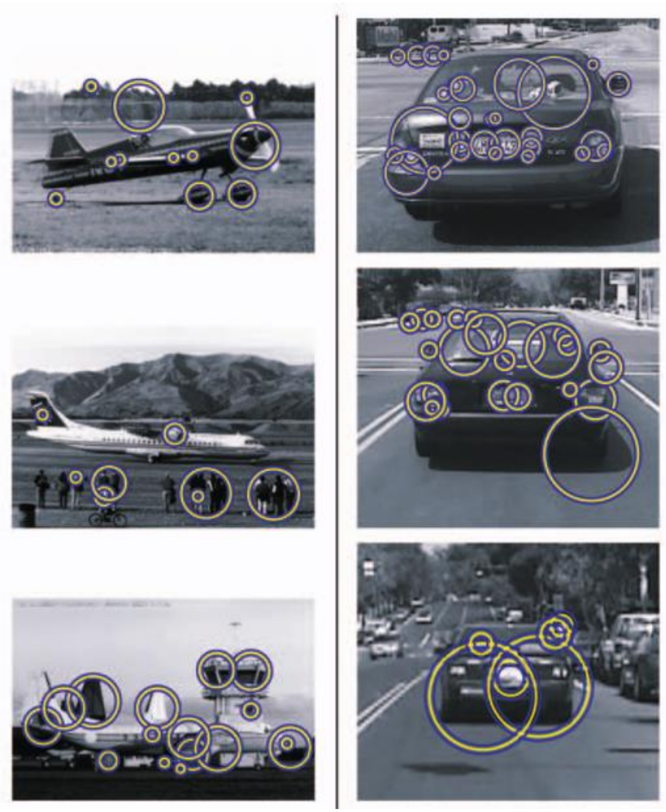


Image from: Y. Chen, J. Bi, and J. Z. Wang, "MILES: Multiple-Instance Learning via Embedded Instance Selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1931–1947, 2006.

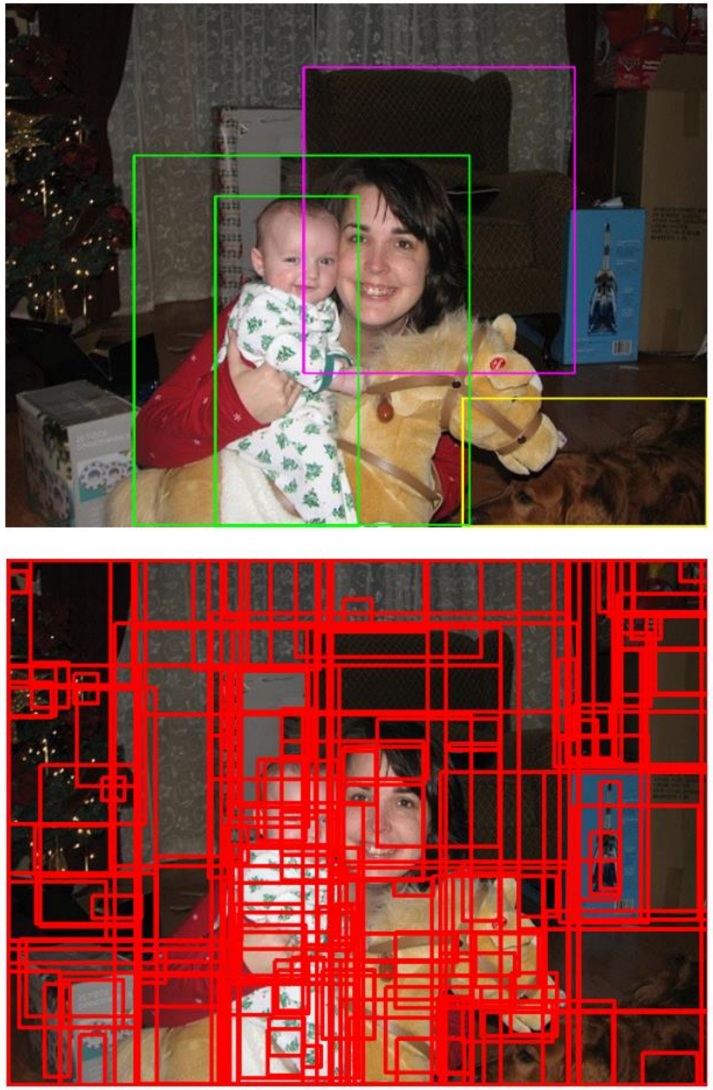
Object Localization in Image

Objective: Find objects in images.

Bag: Collection of candidate annotation boxes

Instance: Sub-image corresponding to candidate windows.

Justification: A large quantity of data can be used to learn because costly strong annotations are not necessary.



H. O. Song, R. Girshick, S. Jegelka, J. Mairal, Z. Harchaoui, and T. Darrell, “On learning to localize objects with minimal supervision,” *International Conference on Machine Learning*, 2014

Computer Aided Diagnosis (from images)

Objective: Predict if a subject is diseased or healthy.

Bags: Collection segments or patches extracted from a medical image.

Instances: Image segments or patches.

Justification: A large quantity of images can be used to train. Only a diagnosis is required per image. Expert local annotation are no longer required.

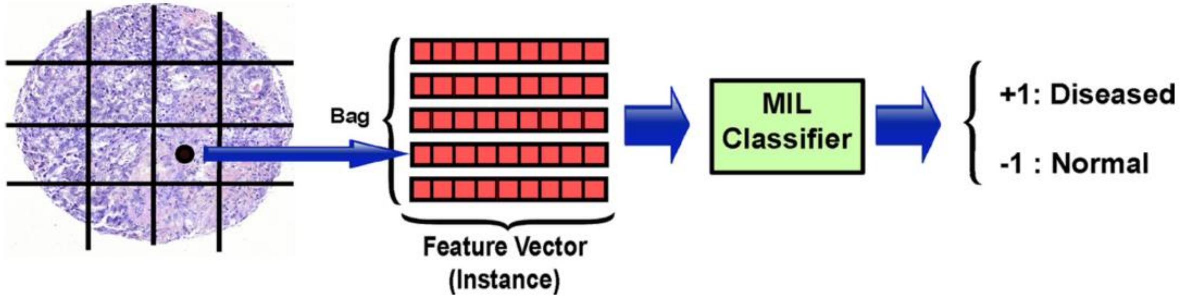


Image from: M. Kandemir and F. A. Hamprecht, “Computer-aided diagnosis from weak supervision: a benchmarking study.,” *Comput. Med. Imaging Graph.*, vol. 42, pp. 44–50, Jun. 2015.

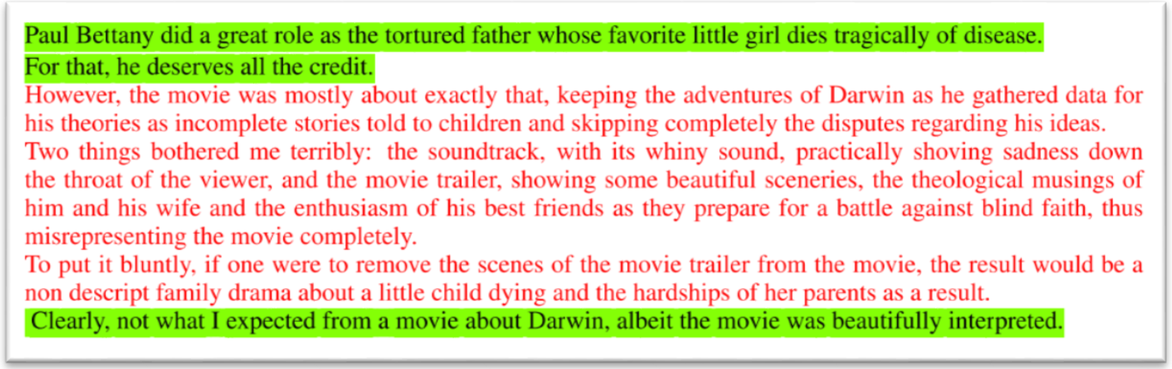
Sentiment Analysis in Text

Objective: Predict if a text/sentence expresses positive or negative sentiment.

Bags: Texts/paragraphs.

Instances: Sentences.

Justification: Large quantity of text can be harvested from the web. A sentiment is usually given to a complete text while it may contain positive and negative sentences.



Paul Bettany did a great role as the tortured father whose favorite little girl dies tragically of disease. For that, he deserves all the credit.
However, the movie was mostly about exactly that, keeping the adventures of Darwin as he gathered data for his theories as incomplete stories told to children and skipping completely the disputes regarding his ideas.
Two things bothered me terribly: the soundtrack, with its whiny sound, practically shoving sadness down the throat of the viewer, and the movie trailer, showing some beautiful sceneries, the theological musings of him and his wife and the enthusiasm of his best friends as they prepare for a battle against blind faith, thus misrepresenting the movie completely.
To put it bluntly, if one were to remove the scenes of the movie trailer from the movie, the result would be a non descript family drama about a little child dying and the hardships of her parents as a result.
Clearly, not what I expected from a movie about Darwin, albeit the movie was beautifully interpreted.

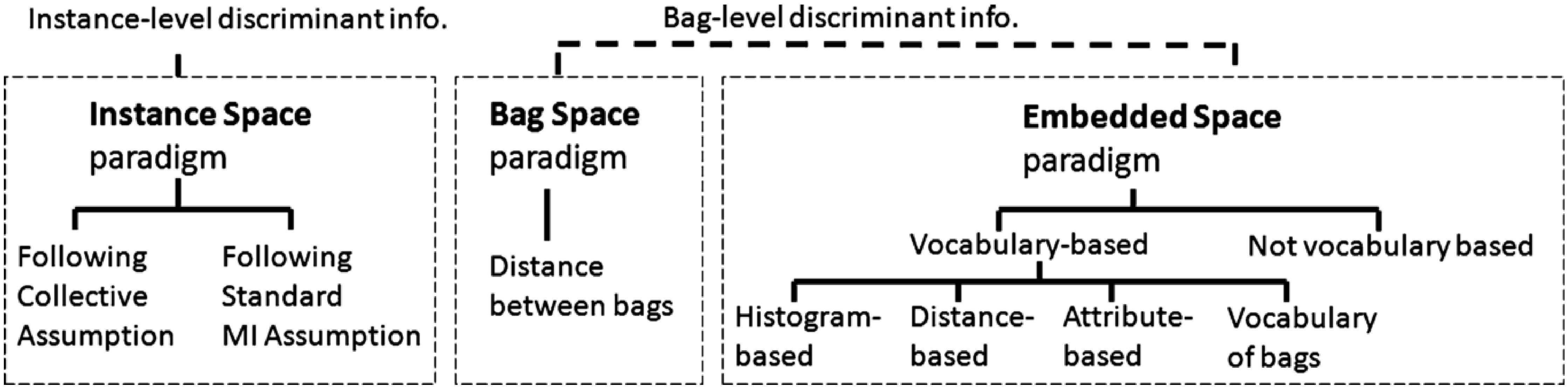
Image from: D. Kotzias, M. Denil, P. Blunsom, and N. de Freitas, “Deep Multi-Instance Transfer Learning,” *CoRR*, vol. abs/1411.3, 2014.

How Can I Do Multiple Instance Learning?

Types of Methods

Taxonomy of MIL Methods

A generally accepted taxonomy divides MIL methods based on their reasoning space:



Taxonomy from: J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, Aug. 2013.

Instance Space Methods

These methods try to uncover the true nature of each instance in order to make a decision on bag labels.

Pros:

- Can be directly used for instance classification tasks.

Cons:

- Do not work when instances have no precise classes.
- Usually less accurate than bag space methods.

MI-SVM
mi-SVM
APR
RSIS
EM-DD
MIL-Boost
SbMIL

Bag Space Methods

These methods embed the content of bags in a single feature vector, thus transforming the problem into supervised learning. Alternatively, they use set distance metrics to compare bags directly.

Pros:

- Can model distributions and relation between instances.
- Deal with unclassifiable instances.
- Can be faster than instance based methods (when an embedding is used).
- Often more accurate for bag classification tasks.

Cons:

- Cannot be directly used for instance classification tasks.

MILES
Citation-kNN
MnID
Bag-of-Words
NSK-SVM
CCE
Mi-Graph
EMD-SVM

What Differentiates Multiple Instance Learning Problems?

Characteristics of MIL problems

Characteristics of MIL Problems

There are characteristics that differentiates a MIL problem from other MIL problems. These characteristics can be related to four distinctive properties of MIL.

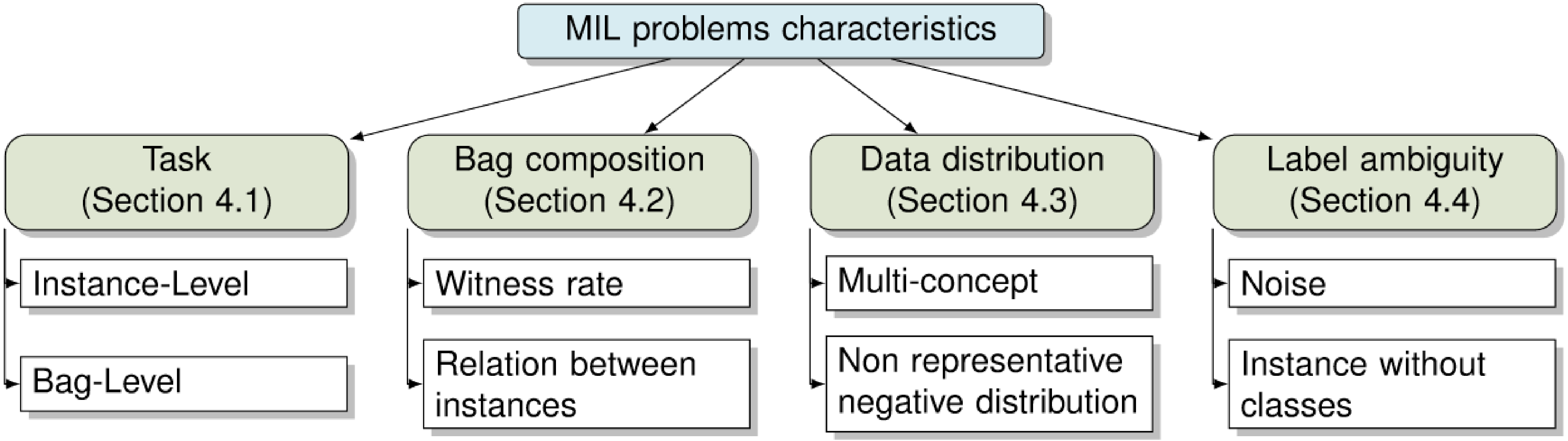


Image from: M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple Instance Learning: A Survey on Problems Characteristics and Applications," to be submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

Task

Instance and bag classification are two different tasks.

It has been observed by many authors that the best algorithm for instance classification is rarely the best for bag classification.

G. Vanwinckelen, V. do O, D. Fierens, and H. Blockeel, "Instance-level accuracy versus bag-level accuracy in multi-instance learning," *Data Mining Knowledge Discovery*, 2015.

The key difference is the instance misclassifying cost.

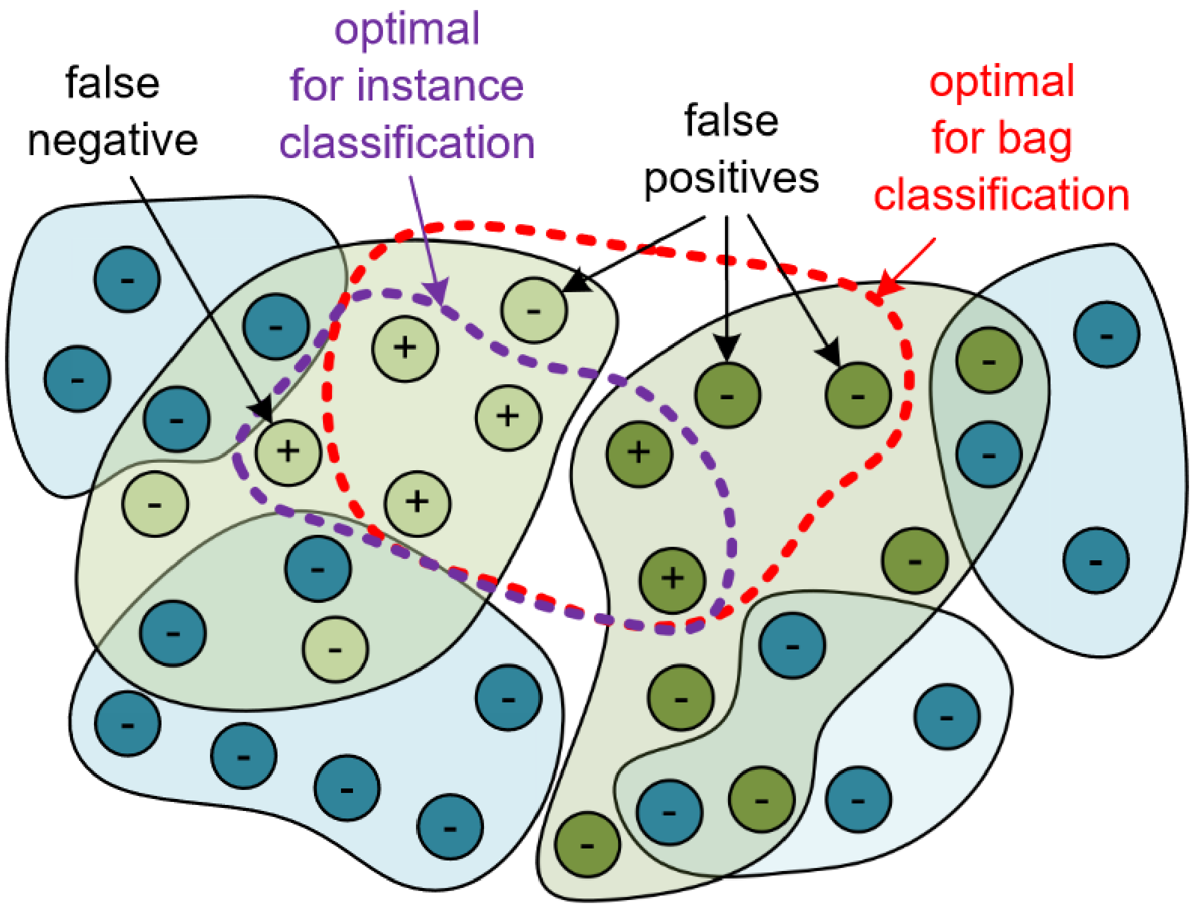


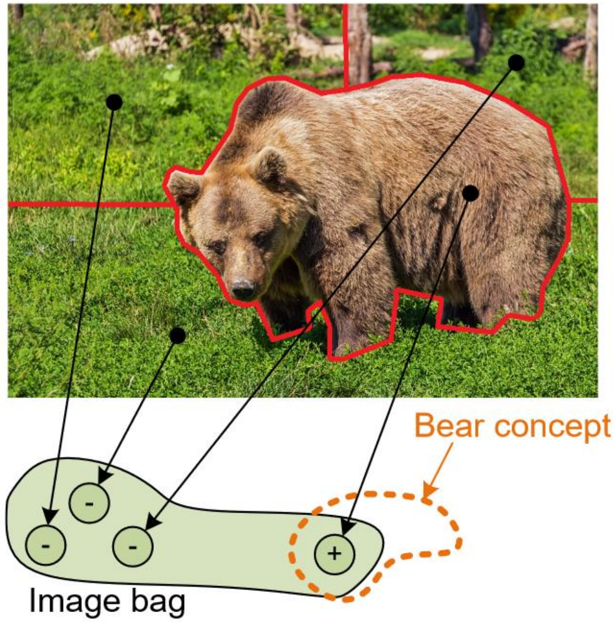
Image from: M.-A. Carbonneau, E. Granger, and G. Gagnon, "Decision Threshold Adjustment Strategies for Increased Accuracy in Multiple Instance Learning," in *Proc. The 6th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2016.

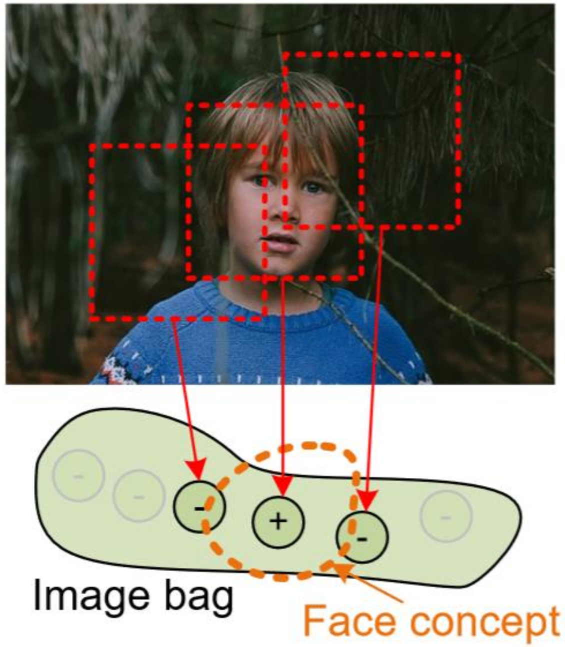
Bag Composition

Depending on the applications, bags can differ in:

- The proportion of positive instances in positive bags (witness rate).
- The size of the bags.
- The relation between the instances:
    - Co-occurences
    - Structure
    - Intra-bags similarities

Images from: M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple Instance Learning: A Survey on Problems Characteristics and Applications," to be submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.



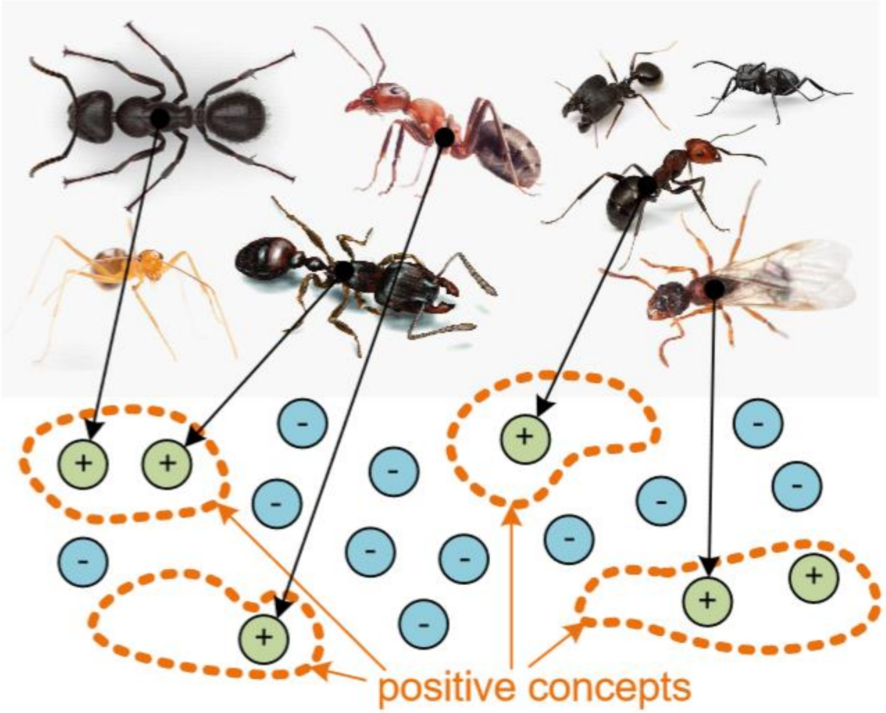


Data Distribution

The type of distribution is important when choosing a MIL algorithm.

Not all MIL algorithms easily deal with :

- Multi-modal distributions
- Unknown negative distribution

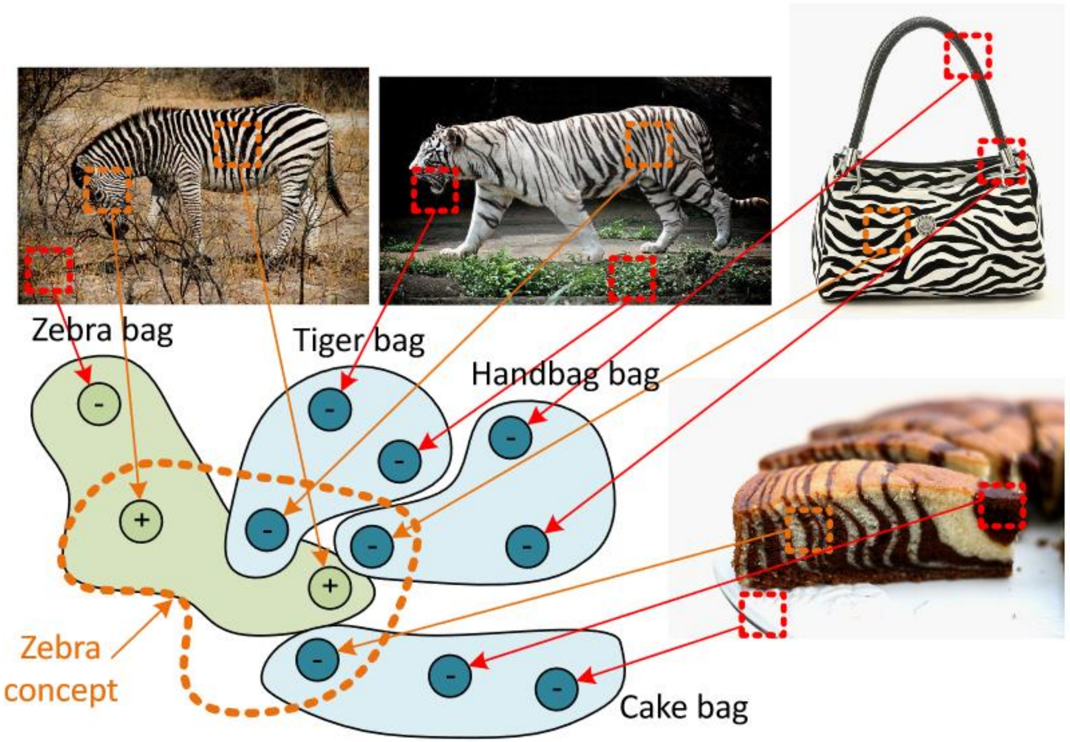


Images from: M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, “Multiple Instance Learning: A Survey on Problems Characteristics and Applications,” to be submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

Label Ambiguity

Weak supervision implies label ambiguity. The ambiguity can be due to:

- Noise.
- Lack of clear classes at instance level.
- Ambiguous representation.
- Classes can share the same type of instances.



Images from: M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple Instance Learning: A Survey on Problems Characteristics and Applications," to be submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

So…

Conclusion

Conclusion

Multiple instance deal with problems where:

- Data points are grouped in sets
- Weak supervision is provided

It is used when:

- Problems are naturally formulated as MIL.
- Strong supervision is costly to obtain or a large quantity of weakly labeled data can be leveraged.

There are several particularities inherent to this type of problem that have to be understood in order to be successful in the application of MIL.