Fundamenta Informaticae XX (2006) 1–15
IOS Press

Learning Sunspot Classification

Trung Thanh Nguyen, Claire P. Willis, Derek J. Paddon
Department of Computer Science, University of Bath
Bath BA2 7AY, United Kingdom

Sinh Hoa Nguyen
Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008, Warsaw, Poland

Hung Son Nguyen
Institute of Mathematics, Warsaw University,
Banacha 2, Warsaw 02-095, Poland

Abstract. Sunspots are the subject of interest to many astronomers and solar physicists. Sunspot observation, analysis and classification form an important part of furthering the knowledge about the Sun. Sunspot classification is a manual and very labor intensive process that could be automated if successfully learned by a machine. This paper presents machine learning approaches to the problem of sunspot classification. The classification scheme attempted was the seven-class Modified Zurich scheme [18]. The data was obtained by processing NASA SOHO/MDI satellite images to extract individual sunspots and their attributes. A series of experiments were performed on the training dataset with an aim of learning sunspot classification and improving prediction accuracy. The experiments involved using decision trees, rough sets, hierarchical clustering and layered learning methods. Sunspots were characterized by their visual properties like size, shape, positions, and were manually classified by comparing extracted sunspots with corresponding active region maps (ARMaps) from the Mees Observatory at the Institute for Astronomy, University of Hawaii.

Keywords: sunspot classification, layered learning, rough sets, machine learning

1. Introduction

Sunspots have been the subject of interest to astronomers and astrophysicists for many years: sunspot sightings were first recorded in China as far back as 165 BC; Galileo made some of the first detailed hand-drawings of sunspots in 1610 using a primitive telescope. With the advent of more sophisticated telescopes and photographic devices, knowledge about sunspots and their relationship to other solar phenomena has increased. Nowadays, it is known that sunspots do not appear to be randomly scattered over the Sun's surface but are confined to a specific band. Sunspots are also recognized to have their own life-cycle. They are born and die, grow and shrink in size, form groups and formations, and move across the Sun's surface throughout their lifetime.

Sunspot observation, analysis and classification form an important part in furthering knowledge about the Sun, the solar weather and its effect on earth [17]. Certain categories of sunspot groups are associated with solar flares. Observatories around the world track all visible sunspots in an effort detect flares at an early stage of their formation. Sunspot recognition and classification are currently manual and labor intensive processes which could be automated if successfully learned by a machine.

Some initial attempts at automated sunspot recognition and classification were presented in [6]. Several learning algorithms were examined to investigate the ability of machine learning in dealing with the problem of sunspot classification. The experiment showed that it is very difficult to learn the classification scheme using only visual properties as attributes. Many characteristics of sunspots cannot be precisely determined from digital images.

To further improve the classification accuracy experiments were performed with classification learning in combination with clustering and layered learning methods. One possible way of improving accuracy is to embed the domain knowledge into the learning process. In previous papers we have considered the case where domain knowledge was given in the form of concept ontology and have presented a rough set and layered learning based method that successfully makes use of such kind of domain knowledge [7] [9]. In our recent paper [1], that approach is applied to the sunspot classification problem with an exception that the concept ontology is *not given* but constructed by a supervised learning method. The proposed solution has been implemented and the experimental results show many advantages in comparison with standard learning algorithms.

2. About Sunspots

2.1. Physical properties

Sunspots are regions in the Sun's photosphere where intense magnetic fields cause the temperature and radiation to be less than in the surrounding, hotter and brighter photosphere gases. A single sunspot consist of one or more dark cores, called *umbrae*, often surrounded by a less dark area called *penumbra*. In the umbrae, very intense, longitudinally oriented magnetic fields cause the photospheric gases to become very cool, and thus dark compared to overall photosphere (see left part of Figure 1).

Sunspots have a tendency to appear in magnetically bi-polar groups. In each group there are normally two major spots, oriented approximately east-west, called the *leading*, preceding or western, and the *following* or eastern spot. The leading spot is usually larger in size and has stronger magnetic field strength. It is first to form, first to develop penumbra, and last to dissipate. Also the leading spot is often located slightly closer to the equator than the following spot.

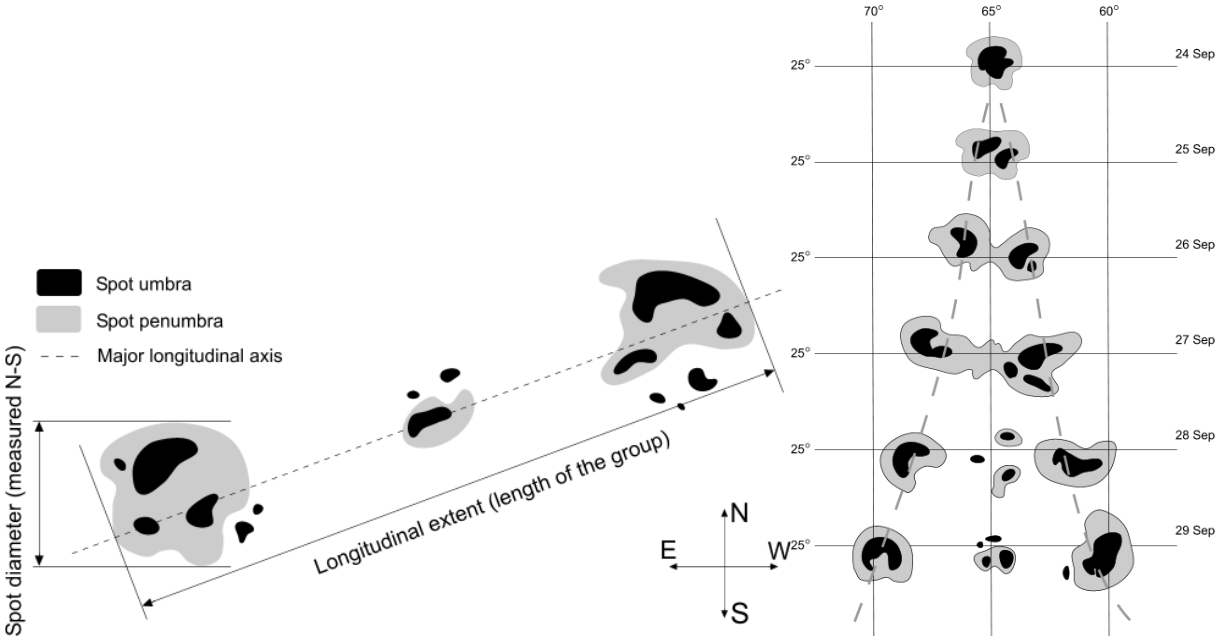

Figure 1. Left: The drawing showing how sunspots diameter and group length should be measured.
Right: The drawing showing that over time differential rotation widens the longitudinal separation between the leading and following spot leading to a class change for the group.

Sunspots exhibit ”proper motion” due to the growth and expansion of the magnetic flux loops in emerging magnetic regions, and differential solar rotation. The polar regions of the Sun rotates slower than the equatorial regions. Since the leading spot usually appears at lower altitude than the following spot over time differential rotation widens the longitudinal separation between these spots (see right part of Figure 1) Once a sunspot has reached its maximum longitudinal extent, it usually stabilizes or starts to decay as the magnetic field weakens. Sunspots within a region will sometimes move relative to each other (e.g. converge or revolve about each other) or the major spot may rotate about an axis.

The number of spots in a sunspot group is the number of umbrae (dark cores) visible. For example, two umbrae surrounded by the same penumbral area count as two spots. Location (latitude and longitude) use the generic center of the group. Length of the group (longitudinal extent) is a measure between the outermost extremities of the groups' leading and the following ends (results are given in heliographic degrees). Major axis is usually not parallel to latitude lines.

2.2. Classification scheme

Sunspots appear on the solar disk as individual spots or as a group of spots. Sunspot groups can have an infinite variety of formations and sizes, ranging from small solo spots to giant groups of spots with complex structure. Despite such a diversity of shapes and sizes astronomers have been able to define broad categories of sunspot groups. Using the McIntosh Sunspot Classification Scheme [18] spots are classified according to three descriptive codes. The first code is a modification of the old Zurich scheme, with seven broad categories:

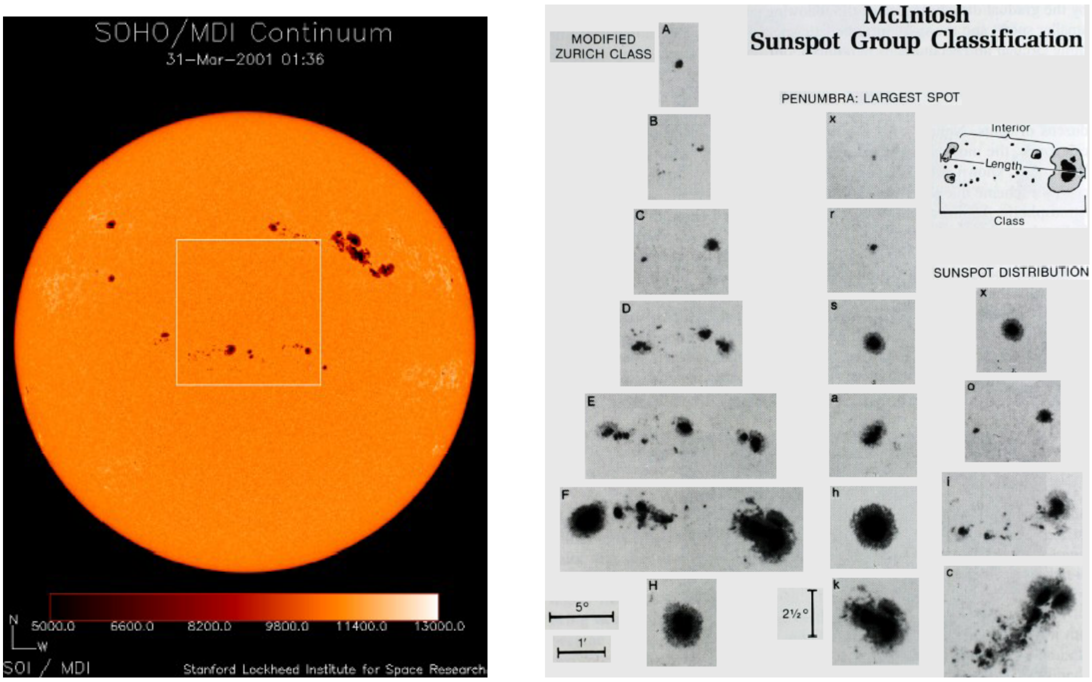

Figure 2. Left: The SOHO/MDI satellite image of the solar disk, showing sunspots. Right: the McIntosh Sunspot Classification Scheme. (Courtesy P.S. McIntosh, NOAA(1990))

- A: Unipolar group with no penumbra, at start or end of spot group's life
- B: Bipolar group with penumbrae on any spots
- C: Bipolar group with penumbra on one end of group, usually surrounding largest of leader umbrae
- D: Bipolar group with penumbrae on spots at both ends of group, and with longitudinal extent less than 10 arc seconds (120 000 km)
- E: Bipolar group with penumbrae on spots at both ends of group, and with longitudinal extent between 10 and 15 arc seconds
- F: Bipolar group with penumbrae on spots at both ends of group, and length more than 15 arc seconds (above 180 000 km)
- H: Unipolar group with penumbra. Principal spot is usually the remnant leader spot of pre-existing bipolar groups

The second code describes the penumbra of the largest spot of the group and the third code describes the compactness of the spots in the intermediate part of the group [18]. Up to sixty classes of spots are covered, although not all code combinations are used. A particular spot or group of spots may go through a number of categories in their lifetime. Solar flares are usually associated with large groups.

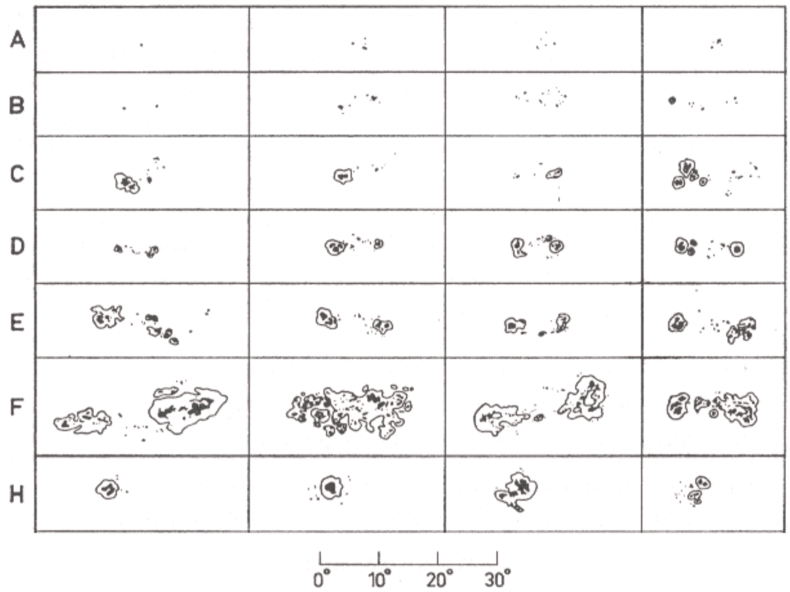

Figure 3. The drawing shows possible visual appearances for each class. There is a wide allowable margin in the interpretation of the classification rules making automatic classification difficult

2.3. Issues with automated learning

When attempting automated classification the following issues need to be considered:

1. **Interpreting classification rules:** As only broad forms of classification exist there is a large allowable margin in the interpretation of classification rules. The same group may be assigned a different class depending on the interpretation of the expert undertaking the classification. Observatories share information and cross-check results regularly to form a consistent opinion.

2. **Individual spots and groups:** Sunspot classification schemes classify sunspot groups not individual spots. When sunspots are extracted from digital images they are treated as individual spots. Hence further information is required to group spots together to form proper sunspot groups.

3. **Dealing with groups migration:** Sunspots have their own life-cycle and migrate across the Sun's surface. They start their life as small tiny spots that usually continue to form pairs and evolve into groups. Once a group attains its maximum size it starts to decay. As a result, a particular group may change its class assignment several times during its lifetime. A reliable method to keep track of those changes must be devised to correctly follow a group during its lifetime. It may be difficult to decide exactly when the change occurs. An individual image of a solar disk containing sunspots has no information about their previous and future class. Moreover, as groups approach the edge of the visible solar disk their shape appears compacted making classification based solely on digital images difficult.

4. **Availability of data:** The average number of visible sunspots varies over time, increasing and decreasing on average over an 11.8 year cycle. As each cycle progresses sunspots gradually start to appear closer to the Sun's equator while forming larger and more complex groups. This creates an issue when deciding on the input data range for a *training dataset*. For example, by taking observations only from a short period at solar maximum, where there are likely to be more sunspots groups class D, E, F, an unbalanced training sample may be obtained.

5. **Quality of input data:** For automatic recognition and classification systems to perform well they need a consistent set of high quality input images, free of distortions and of fairly high resolution. Images should be taken from one source and the same instrument to reduce the variability. Thus satellite images are more suitable than photographs taken from the ground. Note, that some sunspots can be very small and may not be captured at all.

The main difficulty in correctly determining sunspot groups concerns the interpretation of the classification scheme itself. There is a wide allowable margin for each class (see Figure 3). Therefore, classification results may differ between different astronomers doing the classification. In other words, the interpretation of classification rules is said to be non-deterministic.

3. Data collection and preparation

The process of constructing the *training dataset* consisted of gathering data from two sources: the NASA/SOHO website and the ARMaps pages from the Hawaii University website. An ARMap is a map of the solar disk with active regions and sunspots manually classified by expert astronomers [6]. The resulting data set consists of sunspots as objects, their visual properties (size, shape, etc.) as attributes and the Zurich classification (made by experts from ARMaps) as the class label.

Attribute selection: The features extracted by the image processing method were shape descriptors describing the shape of single sunspots and information about spot's neighbors. The following sunspot features were extracted: *x* and *y* coordinates of a spot center; *area* of a spot; *perimeter* length around a spot; spot's *angle* to the main axis; spot's *aspect ratio*, *compactness*, and *form factor*; spot's *feret's diameter*; spot's *circularity*; count of how many neighboring spots are within a specified *radii* (nine radii were selected).

Data preparation: The following manual classification process by an expert astronomer was repeated for all training images: Find an ARMap that fitted the corresponding drawing of detected sunspots using the date and the filename of a drawing. Looked at the regions marked on the ARMap and matched them with the regions of spots detected in the drawings. All regions on the ARMap were numbered - to be annotated. All spots that fell within each identified region were selected. Since each spot is numbered, it was possible to assign the ARMap region number to those spots in the main flat file. All spots with an identical ARMap region number were assigned the class of the ARMap region.

4. Learning methods

Three series of experiments were performed on the same dataset (2589 observations from the period of September 2001 to November 2001). In the first series of experiments (see [6]) we attempted to classify individual spots using decision trees, rough set rules, and an instance-based method (k-nearest neighbor). To improve the results hierarchical clustering was employed in the second series of experiments [2]. Finally we presented how a layered learning approach can improve classification accuracy over the standard learning approach [1].

4.1. Classification learning

In classification learning selecting the right set of attributes for use in the dataset can have a dramatic impact on the performance of the learning scheme and requires an understanding of the problem to be solved through consulting with an expert. Limitations arise from the data source and pre-processing by the image processing module.

The features extracted by the image processing method were mostly shape descriptors describing the shape of single sunspots but containing no information about the spot's neighbors. One way of obtaining such information would be to calculate the distances to the nearest neighbors or to count how many spots are within a certain radius of the target.

For example a spot that is located somewhere inside a group of class F would be expected to have many neighbors. This can be contrasted with a spot of class H that has no immediate neighbors. Moreover within each bipolar group, there are always one or two leading spots, which are substantially larger than the rest of the spots in the group. Moving from class B to F these leading spots gets larger in size. Therefore, for any spot if the *number of neighbors*, within a certain *radius*, and their *sizes* could be determined it would almost certainly be possible to tell which class the spot belongs to.

This meant that the distances between every single spot identified in an image were needed. The value of the radii used to group spots in this experiment were set to reflect 120000 km and 180000 km intervals specified in the Modified Zurich scheme. Radii were set at 60000 km, 120000 km, 180000 km. These values were converted to distances in pixels and scaled. Counts of the number of spots within each radius were computed.

Two data mining tools WEKA [20, 22] and RSES[21, 3] were used which contain learning schemes implemented. The classification "success rate" was determined by the number of *true postives* and *true negatives* over the entire range of classes. This meant that on the resulting confusion matrix high values across the main diagonal line should have been seen.

We applied four well-known classification algorithms on the prepared data set (containing 2589 objects and 20 attributes), namely:

WEKA.J48: The implementation of C4.5[19] decision tree algorithm in WEKA system.

WEKA.Ibk: The implementation of kNN algorithm in WEKA system.

RSES.LEM2: The implementation of LEM2[14] algorithm in RSES system.

RSES.kNN: The RIONA algorithm[13] – the classification algorithm combining rule induction and instance based learning methods. This method is implemented in RSES system.

We then repeated the same experiment but before applying previous classification methods, we selected the most relevant subset of attributes for each learning algorithm. For most algorithms the best subset consisted of attributes describing spots neighborhood and location. Shape descriptors were less relevant. In addition a boosting method, called the *AdaBoostM1* [10], was applied to the J48 algorithm to improve results. Experiment results are summarized in Table 1.

Table 1. Comparison of accuracy and true positive rates of different classification algorithms

Scheme	Accuracy	A	B	C	D	E	F	H
J48 all attributes	73.31 %	0.13	0.33	0.54	0.73	0.73	0.80	0.84
J48 subset	77.33 %	0	0.36	0.60	0.80	0.77	0.83	0.80
J48 subset + boost	85.09 %	0	0.57	0.72	0.88	0.86	0.88	0.81
IBk all attributes	63.89 %	0.25	0.29	0.45	0.66	0.65	0.71	0.54
IBk subset	89.57 %	0.25	0.76	0.85	0.92	0.91	0.94	0.62
RSES kNN all	83.32 %	0.20	0.65	0.72	0.84	0.85	0.86	0.84
RSES kNN subset	90.60 %	0.13	0.59	0.79	0.91	0.94	0.94	0.78
RSES LEM2 all	66.84 %	0.10	0.47	0.46	0.65	0.68	0.72	0.84
RSES LEM2 subset	77.50 %	0	0.55	0.58	0.79	0.80	0.81	0.77

The distribution of classes in our data set is presented in the Table 2

Table 2. The distribution of classes in the dataset

Group classification	A	B	C	D	E	F	H
Class distribution	0.31%	1.62%	7.49%	30.67%	25.45%	28.51%	5.95%

Note that because the dataset itself contained very few examples of class A, B, and C the prediction accuracy for those classes are much lower than the rest, making the overall accuracy figure for each method seem less meaningful. If the dataset is enriched with more examples from those classes we should expect an overall improvement. Moreover, the results show very good accuracy figure for class H, despite having a small population in the dataset. This can be explained that strong rules were found for that class. For a more detailed discussion see [6].

4.2. Clustering

Clustering was used to group individual sunspots together using euclidian distance as a dissimilarity measure. The idea behind clustering is to try to re-create real sunspot groups to improve classification learning results. Experiments were performed with three different hierarchical clustering algorithms: single-link, average-link, and complete-link. The input data was taken from the same dataset constructed for classification learning but pre-labeled with classes. For every day's worth of observations a distance matrix was calculated, and spots clustered using one of the methods.

Since sunspot groups have dimension limits the sum of all spot distances within a cluster was used as a stopping condition. If a diameter of a cluster grows too large the clustering process was stopped. A performance measure used for obtaining the best threshold value was a cluster purity measure. For each cluster produced by the clustering algorithm a comparison was made with a reference cluster to identify how many spots were in fact correctly grouped. A 100% pure cluster is the cluster which had all the spots correctly grouped. Thus to find the best threshold value for the dataset the cluster purity measure was calculated for each cluster and the average obtained for the whole dataset for every threshold value (see Table 3). The objective was to select the method that generates fewer but purer clusters. A comparison of clustering methods is shown in Figure 4.

Table 3. Clustering results

	Single-Link		Average-Link		Complete-Link	
Stop threshold	Overall Purity	Number of clusters	Overall Purity	Number of clusters	Overall Purity	Number of clusters
1000	0.972	886	0.962	702	0.957	671
1500	0.951	752	0.935	593	0.924	557
2000	0.938	661	0.916	538	0.898	494
2500	0.922	599	0.9	482	0.879	451
3000	0.906	544	0.887	450	0.867	419
3500	0.892	498	0.871	413	0.849	389
4000	0.876	467	0.85	386	0.829	364
4500	0.864	438	0.835	363	0.816	342

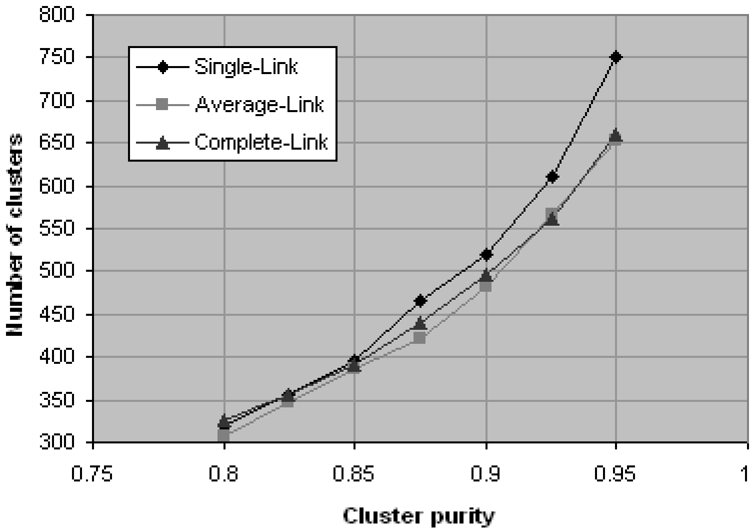

Figure 4. The comparison of three clustering methods: Single-Link, Average-Link and Complete-Link

The single-link method produced more clusters at the same level of cluster purity compared to average-link and complete-link. Whereas average-link and complete-link methods perform almost equally. Ultimately we have chosen to use complete-link method with a stoping threshold value of 2000 to build clusters for layered learning experiments.

4.3. Layered learning

To improve the classification accuracy, we try to embed the domain knowledge into the learning process. Layered learning [4] is an alternative approach to concept approximation. Given a hierarchical concept decomposition, the main idea is to synthesize a target concept gradually from simpler ones. One can imagine the decomposition hierarchy as a treelike structure (or acyclic graph structure) containing the target concept in the root. A learning process is performed through the hierarchy, from leaves to the root, layer by layer. At the lowest layer, basic concepts are approximated using feature values available from a data set. At the next layer more complex concepts are synthesized from basic concepts. This process is repeated for successive layers until the target concept is achieved.

In previous papers (see [8] [7]) we presented a hierarchical learning approach to concept approximation based on rough set theory. The proposition was performed with an assumption that the concept ontology already exists. This assumption is not satisfied in the case of sunspot classification problem. Thus as presented in [1] we constructed the concept decomposition scheme from the domain knowledge. Our solution to the sunspot classification problem using layered learning consists of four main steps:

1. recognize single sunspots using image processing techniques and create decision table describing their classification made by experts;

2. group daily sunspots into clusters and create decision table for those clusters;

3. create a hierarchical decomposition scheme of concepts from domain knowledge;

4. apply hierarchical learning method based on rough set theory to learn the Zurich sunspot classification scheme.

4.4. Construction of the concept ontology

In Section 2.2 we have presented the original sunspot classification scheme. This scheme seems to be complicated but, in fact, the classification can be described by simpler concepts:

1. **Magnetic type of groups:** there are two possible types called *unipolar* and *bipolar*;

2. **Group span:** a heliographical distance of two farthest spots in a group; there are three spanning degrees, i.e., *NULL* (not applicable), *small* (less than 10 h.degs. or 120000 km), *large* (more than 15 h.degs. or 180000 km) and *middle* (between 10 h.degs and 15 h.degs.);

3. **Penumbra type of the leading spot:** there are four possible types called *no penumbra, rudimentary, asymmetric, and symmetric*;

4. **Penumbra size of the leading spot:** there are two possible values *small* (less than 2,5 h.degs. or 30000km), and *large* (more than 2,5 h.degs.);

5. **Distribution of spots inside a group:** there are four possible values called *single, open, intermediate, and compact*.

If we consider all situations described by those five concepts, there are only 60 possible scenarios. Every possibility is characterized by those concepts (which can be treated as attributes) and can be labeled by one of seven letters $\{A, B, C, D, E, F, H\}$, accordingly to the Zurich classification scheme. Therefore we have a decision table with 60 objects, 5 attributes, 7 decision classes. The idea is to create a decision tree for the described above decision table. The resulting tree computed by the decision tree induction method, which is implemented in Weka [22] as J48 classifier, is presented in Figure 5.

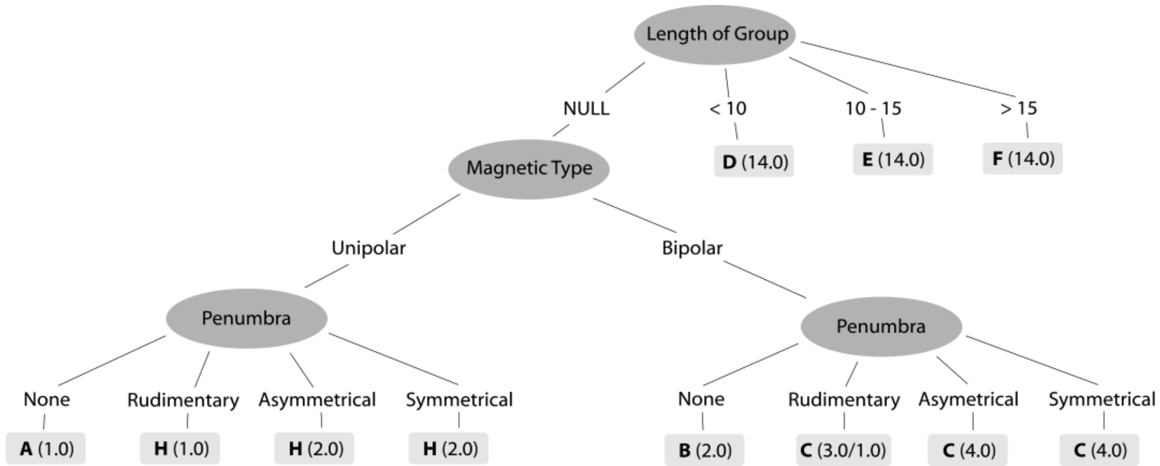

Figure 5. The Zurich classification scheme represented by a decision tree

This decision tree leads to the following observations, which are very useful for concept decomposition process: (1) Classes D, E and F are similar on almost all attributes except attribute **group span**; (2) Classes A, H have similar magnetic type (both are unipolar), but they are discerned by the attribute **penumbra type**; (3) Classes B, C have similar magnetic type (both are bipolar), but they are discerned by the attribute *penumbra size*.

The final concept ontology of the target concept has been build from those observations. Figure 6 presents the main part of this ontology which was created by including the following additional concepts to the decision tree in Fig. 5:

- **Group AHBC?**: does a sunspot cluster belong to one of classes A, B, C, H?
- **Group DEF?**: does a sunspot cluster belong to one of classes D, E, F?
- **AHBC-DEF**: the classification that distinguishes $\{A, B, C, H\}$ and $\{D, E, F\}$
- **A-H-B-C-DEF**: the classification that groups classes D, E, F together;
- **A-H-B-C-D-EF**: the classification that groups classes E, F together;
- **D-EF**, **E-DF**, **F-DE**: classification problems that distinguish one class from the rest for three decision classes D, E, F;

- **target classes**: what is the label of a sunspot cluster?

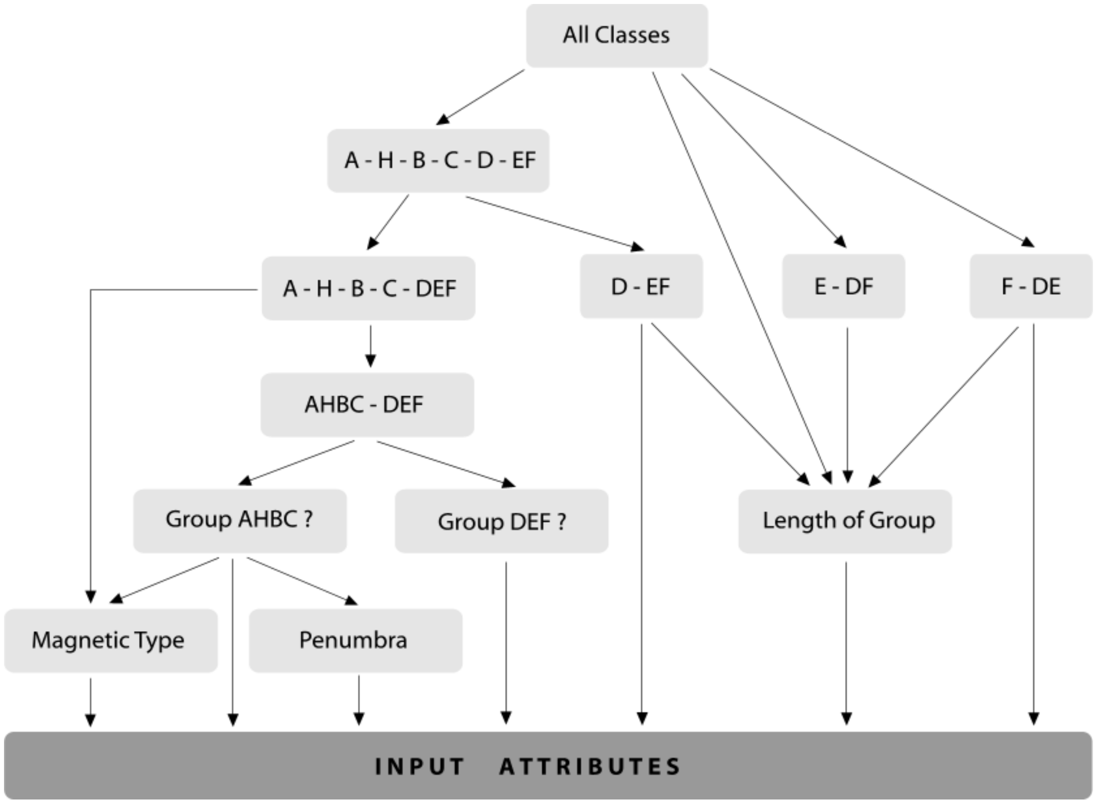

Figure 6. The concept hierarchy for sunspot classification problem

The synthesis process is performed through the concept hierarchy, from leaves to the root (as presented in [7]. The learning algorithm, for every node N of the concept hierarchy, produces the rough membership function for every decision class that occurs in N. Later, the extracted membership functions are used as attributes to construct the rough membership function for those concepts occurring in the next level of the hierarchy.

4.5. Layered learning experiment

For each daily image of solar disk in the investigated period, i.e., from September 2001 to November 2001, we have applied the sunspot recognition algorithm and the clustering algorithm to extracted sunspots. We have obtained, all together, 494 sunspot clusters. The training set (obtained from September and October 2001) consisted of 366 clusters, while the test set (November 2001) contained 128 sunspot clusters. The distribution of decision classes in training and test data is presented in Table 4 and a comparison of standard and layered learning method is shown in Figure 7. For a more detailed discussion of results see [1].

Table 4. The distribution of decision classes on training and test data sets.

Table	No. of obj.	Modified-Zurich classes						
		A	B	C	D	E	F	H
Train set	366	0,8%	2,2%	9,6%	30,6%	19,7%	21,9%	15,3%
Test set	128	0%	1,6%	7,8%	36,7%	18,8%	18%	17,2%

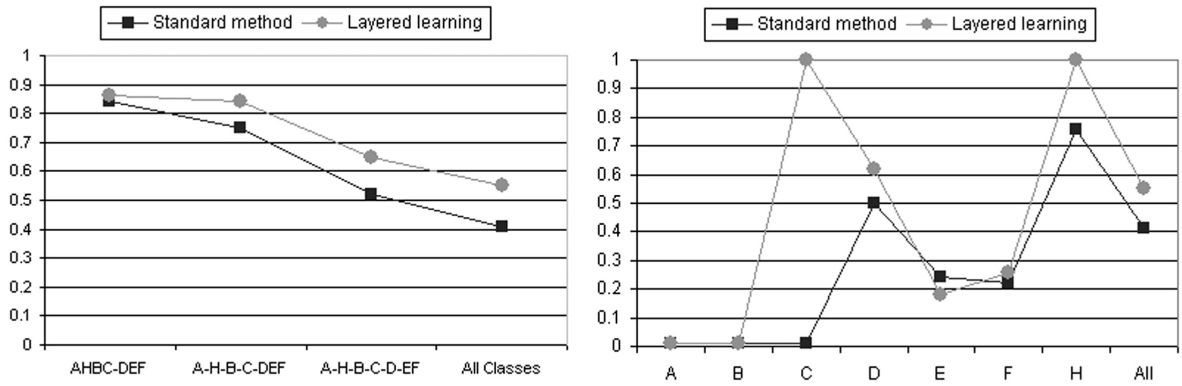

Figure 7. Left: the classification accuracy of standard and layered method for some concepts in the ontology presented in Fig. 6. Right: the classification accuracy of standard and layered method for particular decision classes.

4.6. Summary of results

A considerable improvement was obtained by applying the method based on rough sets and layered learning approach. The highest improvements are achieved for classes C and H which are recognized by layered learning method with 100% accuracy (see Fig. 7). Classes A and B are too small to be evaluated. Also, the accuracy of recognition problem: "whether a cluster belongs to one of three classes D, E, F" is very high (about 98%). The problem remains to separate those three classes. Unfortunately, our clustering algorithm tends to form smaller groups compared to reference ones. Therefore some large clusters were broken into a few smaller ones, and this may be the cause of low classification accuracy.

4.7. Design of a learning system

Sunspot observation and classification is currently done manually by astronomical observatories (ie. NASA/NOAA). The process involves many steps and is very time consuming. Moreover it is subjected to human errors due to the non-deterministic nature of the classification scheme used. Therefore to effectively help astronomers a system needs to be devised that performs the entire workflow, from data capture to classification and cataloguing.

A typical automated sunspot classification system may consist of two modules: the image processing module and the classification module. The aim of the former is to handle the input image, extracting spots and their properties. The classification module is responsible for predicting the spot's class and grouping them together.

Our current experimental system is able to import digital images of solar disks from online NASA SOHO/MDI satellite, separate individual spots from their background using a custom threshold function and extract their features to a text file to build a matrix of instances and attributes. Such a flat-file can be imported to machine learning tools (such as WEKA, RSES) for building a classifier. A future objective would be to build a complete system whose input is an image and output are sunspot groups marked and classified. Such a system would also feature user feedback, trainable classifiers and training data generator.

5. Conclusion

In this paper we presented machine learning approaches to the problem of sunspot classification. Results obtained from classification learning experiments have shown that it is possible to accurately classify individual sunspots using decision trees and rough sets. The results can be further improved if individual spots are clustered and the layered learning method employed. However there are several problems that need solving. Firstly, current dataset need to be enriched and balanced with new examples. The clustering algorithm could be improved to produce groups that more closely match real sunspot groups (ie. fewer groups and purer). Such algorithms as density based clustering are the likely candidates. Finally, in order to improve the layered learning algorithm's performance it may be necessary to change the image processing module to extract more detail from input images. This is especially crucial in the ability to distinguish between classes D, E, and F.

Acknowledgement:

The research has been partially supported by the grant 3T11C00226 from Ministry of Scientific Research and Information Technology of the Republic of Poland and the research grant of Polish-Japanese Institute of Information Technology.

References

[1] Sinh Hoa Nguyen, Trung Thanh Nguyen, Hung Son Nguyen Rough Set Approach to Sunspot Classification Problem In Dominik Slezak, Jingtao Yao, James F. Peters, Wojciech Ziarko, Xiaohua Hu (Eds.): *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, 10th International Conference, RSFDGrC 2005, Regina, Canada, August 31 - September 3, 2005*, Proceedings, Part II. Lecture Notes in Computer Science 3642 Springer 2005. pages 263-272

[2] Trung Thanh Nguyen, Claire P. Willis, Derek J. Paddon, Sinh Hoa Nguyen, and Hung Son Nguyen. Data Mining Approach to Sunspot Classification Problem in *Rough Set Techniques in Knowledge Discovery Workshop at The 9th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-05) in Hanoi, Vietnam 18-20 May 2005.*

[3] Bazan J., Szczuka M. RSES and RSESlib - A Collection of Tools for Rough Set Computations, Proc. of RSCTC'2000, LNAI 2005, Springer Verlag, Berlin, 2001

[4] P. Stone. *Layered Learning in Multi-Agent Systems: A Winning Approach to Robotic Soccer*. The MIT Press, Cambridge, MA, 2000.

[5] P. H. Scherrer, et al., Sol. Phys., 162, 129, 1995.

[6] Trung Thanh Nguyen, Claire P. Willis, Derek J. Paddon, and Hung Son Nguyen. On learning of sunspot classification. In Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, and Krzysztof Trojanowski, editors, *Intelligent Information Systems, Proceedings of IIPWM'04, May 17-20, 2004, Zakopane, Poland*, Advances in Soft Computing, pages 59–68. Springer, 2004.

[7] Sinh Hoa Nguyen, Jan Bazan, Andrzej Skowron, and Hung Son Nguyen. Layered learning for concept synthesis. In Jim F. Peters, Andrzej Skowron, Jerzy W. Grzymala-Busse, Bozena Kostek, Roman W. Swiniarski, and Marcin S. Szczuka, editors, *Transactions on Rough Sets I*, volume LNCS 3100 of *Lecture Notes on Computer Science*, pages 187–208. Springer, 2004.

[8] Sinh Hoa Nguyen and Hung Son Nguyen. Rough set approach to approximation of concepts from taxonomy. In *Proceedings of Knowledge Discovery and Ontologies Workshop (KDO-04) at ECML/PKDD 2004, September 24, 2004, Pisa, Italy*, 2004.

[9] Sinh Hoa Nguyen and Hung Son Nguyen. Learning concept approximation from uncertain decision tables. In *Monitoring, Security, and Rescue Techniques in Multiagent Systems Dunin-Keplicz, B.; Jankowski, A.; Skowron, A.; Szczuka, M. (Eds.), Advances in Soft Computing*, Springer-Verlag 2005, page 249–260.

[10] Freund, Y., and R. E. Schapire *Experiments with a new boosting algorithm*. Proc. Thirteenth International Conference on Machine Learning,Morgan Kaufmann, 1996, pages 148-156.

[11] R. J. Bray and R. E. Loughhead. *Sunspots*. Dover Publications, New York, 1964.

[12] P. Hadjinian R. Stadler J. Verhees Cabena, P. and A. Zanasi. *Discovering data mining: From concept to implementation*. Prentice Hall, Upper Saddle River, NJ., 1998.

[13] G. Gora and A. Wojna., RIONA: A New Classification System Combining Rule Induction and Instance-Based Learning, Fundamenta Informaticae, 51(4), 2002, pages 369–390

[14] Grzymała-Busse J., A New Version of the Rule Induction System LERS Fundamenta Informaticae, Vol. 31(1), 1997, pp. 27–39

[15] R. Kohavi and F. Provost. Machine learning: Special issue on application of machine learning and the knowledge discovery process. *Machine Learning*, 30, 1998.

[16] P. Langley and H. A. Simon. Applications of machine learning and rule induction. *Communications of the ACM*, 38(11):55–64, 1995.

[17] K. J. H. Phillips. *Guide to the Sun*. Cambridge University Press, 1992.

[18] P. McIntosh, Solar Physics 125, 251, 1990.

[19] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[20] I. H. Witten and Frank E. *Data Mining: practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann Publishers, San Francisco, CA., 2000.

[21] The RSES Homepage, `http://logic.mimuw.edu.pl/∼rses`

[22] The WEKA Homepage, `http://www.cs.waikato.ac.nz`