IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 2, No 2, March 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

On Extracting Important Correlations in Salmon Lice Based Dataset

Ingunn Alne Hoell and Sylvia Encheva

Faculty of Technology, Business and Maritime Sciences,Stord/Haugesund University College
Haugesund, 5528, Norway

Abstract

Salmon lice causes reduced fish welfare as well as great economical losses along with a significantly increased amount of work for fish farmers. Resistance among salmon lice to well established treatments has resulted in repeated treatments and reduced immune systems among the fish. The main focus of this work is to investigate correlations between salmon lice- and treatment-data received from a particular area on the Norwegian coast. Data mining originated approaches are used in an attempt to obtain new knowledge from the investigated data sets.

Keywords: *Salmon, Data Mining, Hasse Diagrams, Association Rules*

1. Introduction

Farming of Atlantic salmon most commonly takes place in marine net cages, where the volumes range from 20.000 - 80.000 m^3. The cages may contain as many as 200.000 - 400.000 individuals [8], and are usually located in semi-sheltered coastal bays. One of the greatest problems emerging with the salmon-farming industry is the increase of salmon lice, *Lepeophtheirus salmonis*, closely related to a fundamental change in number of hosts for the parasite. Salmon lice causes both great economical losses and an increased amount of work for fish farmers, as well as health problems for the fish.

L. salmonis are ectoparasitic copepods causing disease in both farmed and wild salmon species, ranging from mild skin damage to stress induced mortality. In addition to the cost associated with treatment against *L. salmonis*, considerable losses are caused by lower classification ratings of slaughtered fish, and reduced growth rate due to a decreased feed intake induced by salmon lice infections. The wounds caused by sea lice expose fish to osmotic stress and to secondary infections, e.g. the higher number of pancreas disease caused by the PD-virus may be explained by this [9]. Resistance among salmon lice to the established treatments have resulted in repeated treatments and reduced immune systems among the fish.

In this paper we discuss salmon lice- and treatment-data received from The Norwegian Animal Health Authorities (AHA) in one zone on the Norwegian coast. This zone is chosen since it has been particularly troubled with salmon lice during the time interval we have condacted our research. The zone has quite a number of fish farms and therefore sufficiently representative with respect to correlations between salmon lice- and treatment data.

2. Background

L. salmonis mate on their host, and the females carry the fertilized eggs in a pair of egg strings containing from 100 to 1000 eggs. When the eggs hatch, the larvae undergo several life stages on their way to adulthood, see Fig. 1. The first three (Nauplius I, II and Copepoditt) are free-living and planctonic stages, where the larvae can spread over large areas. Chalimus in I-IV life stages are attached to the host's skin through a special frontal filament, [10]. During the preadult I and II stages lice move freely over host's skin to feed, where the host's head and back are especially exposed. Finally, *L. salmonis* reach adulthood, which is also their reproductive life stage.

Since farmed fish are present also during winter, this permits adult female lice to produce larvae all year through. The system with net cages also allows release of salmon lice to the surrounding marine environment, and concequently dispersal of salmon lice to both wild and farmed fish. The *L. salmonis* parasite is considered to be a serious threat to the wild salmon and sea trout (*Salmo trutta L.*) populations, [2] and [13].

Abundance of salmon lice on farmed fish may vary during a season and between years. Usually there are more salmon lice present during the winter months than during summer.

IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 2, No 2, March 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

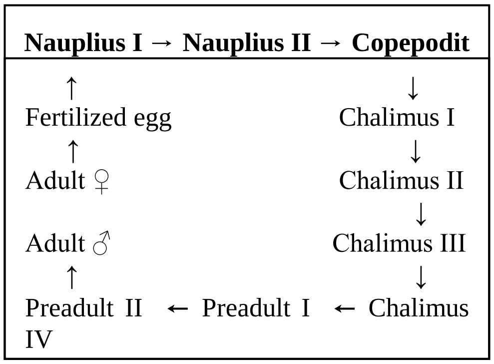

Fig. 1 Schematic presentation of the different life stages of the sea lice *Lepeophtheirus salmonis*. The life stages shown in bold are all free-living and planctonic, life stages (Chalimus I-IV, preadult I-II and adult male and female) are all on the skin of the host fish.

Treatments in Norway should be performed if more than 0.5 adult female salmon lice or more than 3 mobile salmon lice are detected on average per fish in the period between 1^{st} of January and 31^{st} of August. In the period between 1^{st} of September and 31^{st} of December the numbers are 1 adult female salmon lice and 5 mobile salmon lice on average per fish. Treatment against the salmon lice should usually be finished no later than two weeks after exceeding the limits described above.

Table 1: Overview of the chemotherapeutics used to treat salmon against salmon lice in Norway

Chemical group	**Common name**	**Medical agent**	**Treatment method**	***L. salmonis*** **life stages the agent is effective on**
Pyretroid	Alpha max	Deltametrine	Bath	Adult and pre-adult
Pyretroid	Betamax vet.	Cypermetrin	Bath	Adult and pre-adult
Organophos-phate	Salmosan vet.	Azamethiphos	Bath	Adult and pre-adult
Oxidative agent	Hydrogen peroxide	H_2O_2	Bath	Adult and mobile life stages
Avermectin	Slice	Emamectin benzoate	Oral/feed	All
Chitin synthase inhibitor	Releeze vet.	Diflubenzuron	Oral/feed	None mature life stages
Chitin synthase inhibitor	Ektobahn vet.	Teflubenzuron	Oral/feed	None mature life stages

Effective management and control of salmon lice requires good routines and sometimes treatment with antiparasitic compounds. Current treatment methods still rely mostly on chemotherapeutics, optimization the time of treatment, and regional delousing, but e.g. biological control using wrasses (fish from the *Labridae* family, in particular *Labrus bergylta*, *Ctenolabrus rupestris*, and *Symphodus melops*) are also more frequently used these days. The most commonly used and approved chemotherapeutics in Norway can be divided into bath treatments and oral (feed) treatments, see Table 1.

One of the greatest problems associated with use of chemotherapeutics is related to development of resistance among salmon lice. Resistance can be buid up when a population repeatedly is treated with too little chemotherapeutics. This often results in a survival of a few salmon lice. The latter tolerate more chemotherapeutics, and genes encoding this tolerance spread in the population. An important approach to avoid resistance development is to apply treatments based on different medical agents.

3. Methodology

3.1 Concept of a context

Definition 1 *[4] Let P be a non-empty ordered set. If $sup\{x, y\}$ and $inf\{x, y\}$ exist for all $x, y \in P$, then P is called a lattice.*

In a lattice illustrating partial ordering of knowledge values, the logical conjunction is identified with the meet operation and the logical disjunction with the join operation.

Definition 2 *[15] A context is a triple (G, M, I) where G and M are sets and $I \subset G \times M$. The elements of G and M are called objects and attributes respectively.*

For $A \subseteq G$ and $B \subseteq M$, define

$$A' = \{m \in M \mid (\forall g \in A)\ gIm\},$$
$$B' = \{g \in G \mid (\forall m \in B)\ gIm\}$$

where A' is the set of attributes common to all the objects in A and B' is the set of objects possessing the attributes in B.

Definition 3 *[15] A concept of the context (G, M, I) is defined to be a pair (A, B) where $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$.*

The *extent* of the concept (A,B) is A while its *intent* is B. A subset A of G is the extent of some concept if and only if $A'' = A$ in which case the unique concept of the which A is an extent is (A, A'). The corresponding statement applies to those subsets $B \in M$ which is the intent of some concepts.

3.2 Frequent Sets

Frequent sets are sets of attributes that occur often enough to deserve further consideration, [1].

Definition 4 *[3] An association rule $Q \rightarrow R$ holds if there are sufficient objects possessing both Q and R and if there are sufficient objects among those with Q which also possess R.*

The complexity of mining frequent itemsets is exponential and algorithms for finding such sets have been developed by many authors such as [1] and [3].

A context (G, M, I) satisfies the association rule $Q \rightarrow R_{minsup,minconf}$, with $Q, R \in M$, if

$$sup(Q \rightarrow R) = \frac{|(Q \cup R)'|}{|G|} \geq minsup,$$

$$conf(Q \rightarrow R) = \frac{|(Q \cup R)'|}{|Q'|} \geq minconf$$

provided *minsup* $\in [0,1]$ and *minconf* $\in [0,1]$.

The ratios

$$\frac{|(Q \cup R)'|}{|G|} \text{ and } \frac{|(Q \cup R)'|}{|Q'|}$$

are called, respectively, the *support* and the *confidence* of the rule $Q \rightarrow R$. In other words the rule $Q \rightarrow R$ has support $\sigma\%$ in the transaction set T if $\sigma\%$ of the transactions in T contain $Q \cup R$. The rule has confidence $\psi\%$ if $\psi\%$ of the transactions in T that contain Q also contain R.

Algorithms for fast discovery of association rules have been presented in [5], [6], [7], [11], [12], and [14].

4. Data Analysis

Dependences among mobile lice and treatments are shown in Fig. 2, while Fig. 3 points to dependences among adult lice and treatments in the discussed area. The two concept lattices are used to illustrate structures generated from the given data set.

Table 2: Locations, weeks and treatments

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	p	q	r	s	t
O1				×	×	×	×				×				×				
O2		×		×	×		×				×				×				
O3		×		×	×	×	×				×				×				
O4	×											×		×					
O5					×	×	×	×	×	×	×				×				×
O6						×	×	×	×						×				×
O7	×											×			×				×
O8	×			×	×	×							×						
O9	×															×	×		
O10																		×	
O11						×	×	×				×						×	

Table 2 is usually reffered to as an information table, where dependences between objects and their attributes are clearly shown.

Abbreviations used in Table 2, Fig. 2 and Fig. 3 are as follows:

O1- 1st location
O2- 2nd location
…
O11 - 11th location
a - 2nd week
b - 4th week
c - 6th week
…
j - 20th week
k - salmosan applied once
l - salmosan applied two times
m - salmosan applied three times
n - amx (alpha max) applied once

o - H_2O_2 applied once
p - H_2O_2 applied two times
q - slice applied once
r - bmx (betamax vet.) applied once
s - komb. (combination) applied once
t - komb. applied two times

Hasse diagrams based on data from Table 2 are depected in Fig. 2 and Fig. 3. They are both rotated in order to fit the required style. A node in a concept lattice (known also as Galois lattice) shows objects possesing attributes. Lower nodes contain more objects and less attributes. Upper nodes contain less objects which logically enough share more attributes.

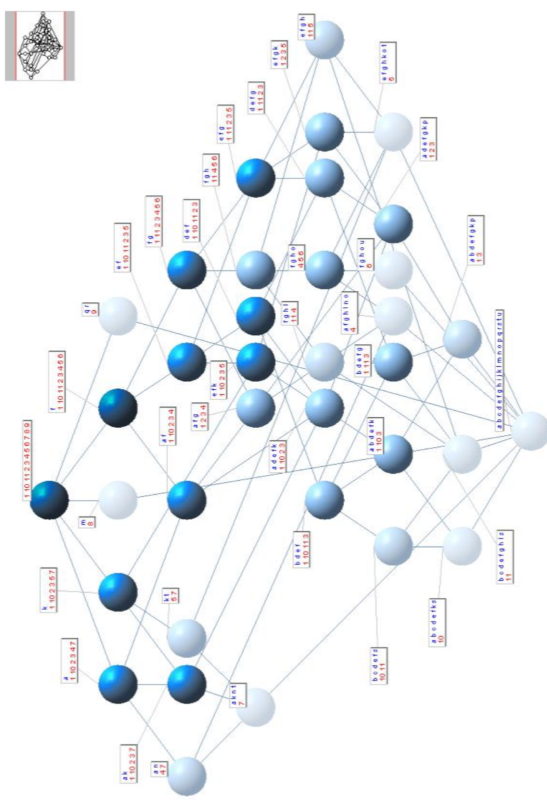

Fig. 2: Correlations among mobile lice and treatments in one are

Some of the most interesting association rules $Q \rightarrow R$ derived from the relationships in Table 1 are coming next.

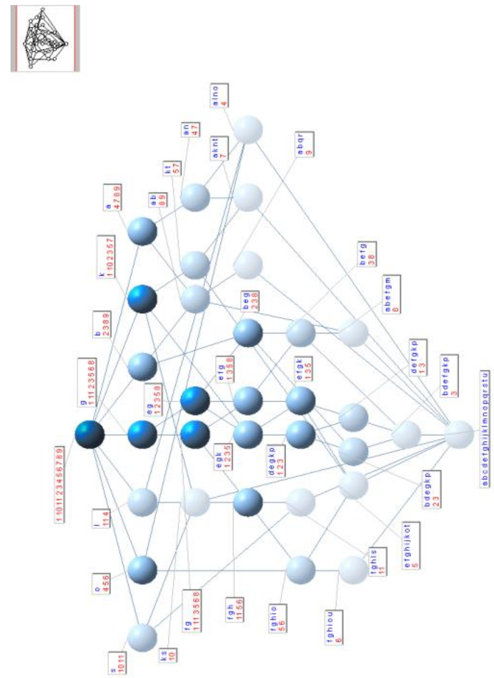

Fig. 3: Correlations among adult female lice and treatments in one area

Association rules concerning adult lice

- Q : n ; R : a
support $Q \rightarrow R$ = 18% , *confidence* $Q \rightarrow R$ =100%

This means that amx was applied once in 18% of the locations, while in all of the locations where amx was applied once the number of adult lice above the average was registered in the second week.

- Q : k ; R : t
support $Q \rightarrow R$ = 18% , *confidence* $Q \rightarrow R$ =33%

This means that salmosan was applied once in 18% of the locations, while in 33% of the locations where salmosan was applied once and komb. was applied two times.

- Q : p ; R : d, e, g, k
support $Q \rightarrow R$ = 27% , *confidence* $Q \rightarrow R$ =100%

This means that H_2O_2 was applied two times in 27% of the locations, while in all of the locations where H_2O_2 was applied two times, salmosan was applied once and the number of adult lice above the average was registered in weeks 8, 10, and 14.

- Q : b, k ; R : d, e, g, p

support $Q \rightarrow R$ = 18% , *confidence* $Q \rightarrow R$ =100%

This means that single application of salmosan and number of adult lice above the average in 4th week was registered in 18% of the locations, while in all of these locations H_2O_2 was applied two times and the number of adult lice above the average was registered in weeks 8, 10, and 14.

Association rules concerning mobile lice.

- Q : a ; R : k

support $Q \rightarrow R$ = 45%, *confidence* $Q \rightarrow R$ =83%

This means that the number of mobile lice above the average in 2nd week was registered in 45% of the locations, while in 83% of these locations salmosan was applied once.

- Q : k ; R : t

support $Q \rightarrow R$ = 18% , *confidence* $Q \rightarrow R$ =33%

This means that salmosan was applied once in 18% of the locations, while in 33% of the locations where salmosan was applied once and komb. was applied two times.

- Q : a, g ; R : d, e, k, p

support $Q \rightarrow R$ = 27% , *confidence* $Q \rightarrow R$ = 75%

This means that the number of mobile lice above the average in weeks 2 and 12 was registered in 27% of the locations, where the number of mobile lice above the average in 8th and 10th weeks was registered in 75% of these locations with mobile lice above the average in 2nd and 12th weeks. Salmosan was applied once and H_2O_2 was applied two times in these locations. The number of mobile lice was above the average in 8th and 10th weeks in the same locations.

- Q : p : R : a, d, e, f, g, k

support $Q \rightarrow R$ = 27%, *confidence* $Q \rightarrow R$ = 100%

This means that H_2O_2 was applied two times in 27% of the locations, while in all of these locations salmosan was applied once and the number of mobile lice above the average was registered in weeks 2, 8, 10, 12, and 14.

- Q : b, g ; R : d, e, f

support $Q \rightarrow R$ = 27% , *confidence* $Q \rightarrow R$ = 100%

This means that the number of mobile lice above the average in weeks 4 and 14 was registered in 27% of the locations, while in all of these locations the number of mobile lice above the average was also registered in weeks 8, 10, and 12.

- Q : b, p ; R : a, d, e, f, g, k

support $Q \rightarrow R$ = 18% , *confidence* $Q \rightarrow R$ = 100%

This means that the number of mobile lice above the average in the 4th week and application of H_2O_2 two times was registered in 18% of the locations, while in all these locations salmosan was applied once and the number of mobile lice above the average was also registered in weeks 2, 8, 10, 12, and 14.

Discussion

The main goal in this paper was to unveil dependencies between salmon lice numbers and different treatments in a chosen zone. To do so, the treatment threshold was used to see if the availability of sea lice deminished completely or their number went below the maximum weakly value for each location. These data were combined with treatments data at each location. Since there is no treatment threshold for attached lice, we have not discussed it. Instead, we focus our study on the number of adult female lice and on mobile lice data. The salmon breeding zone analyzed in this paper was chosen because it was particularly troubled with salmon lice during the period of our investigations. This zone consists of several separate locations, whereof eleven of them are analyzed in further details here. These locations were chosen simply due to their complete dataset reported to AHA for that particular period of time. Ten of them were over the treatment threshold during some parts of the period. None of them were over the treatment threshold for the whole period, but as many as 27 % of the locations were over the treatment threshold for 40 % of the period. These numbers illustrate the huge problems with the salmon lice for the fish farms in this zone at the time of this study.

5. Conclusion

Our data analysis resulted in several association rules, which emphisize important relationships between the salmon lice numbers and the given treatments. Similar relationships can be difficult to observe without conducting such studies. We believe this can provide salmon breeding industry with new approaches for analyzing treatments' effects on salmon lice.

Acknowledgments

We are grateful to Cecilie Ihle at Mattilsynet, Norway for helpful discussions.

References

[1] T. Bastide, R. Taouil, N. Pasquier, G. Stumme, and L. Lakhal,"Mining frequent patterns with counting inference. SIGKDD explorations", Special issue on scalable algorithms, 2000, Vol. 2(2), pp 71–80

[2] P. A. Bjørn, B. Finstad, and R. Kristoffersen, "Salmon lice infection of wild sea trout and Arctic char in marine and freshwaters: the effects of salmon farms", Aquaculture Research, 2001, Vol. 32, pp. 947-962

[3] C. Carpineto, and G. Romano, Concept Data Analysis: Theory and Applications, John Wiley and Sons, Ltd., 2004

[4] B. A. Davey and H. A. Priestley, Introduction to lattices and order. Cambridge University Press, Cambridge, 2005

[5] Z. H. Deng and Z. Wang, "A New Fast Vertical Method for Mining Frequent Patterns", International Journal of Computational Intelligence Systems, 2010, vol. 3(6), pp. 733 – 744

[6] Z. H. Deng, Z. Wang, and J. Jiang, "A New Algorithm for Fast Mining Frequent Itemsets Using N-Lists", Science China Information Sciences, 2012, Vol. 55(9), pp.2008 - 2030

[7] P. Hajek, T. Feglar, J. Rauch, and D. Coufal, The GUHA method, data preprocessing and mining, Database Support for Data Mining Applications, Springer, 2004

[8] F. Oppedal, T. Dempster, and L. H. Stien, "Environmental drivers of Atlantic salmon behaviour in sea-cages: A review", Aquaculture, 2011, Vol. 311, pp. 1-18

[9] E. Petterson, M. Sandberg, and N. Santi, "Salmonid alphavirus associated with Lepeophtheirus salmonis (Copepoda: Caligidae) from Atlantic salmon", Salmo salar L. Journal of Fish Diseases, 2009, Vol. 32, pp. 477-479

[10] A. W. Pike, and S. L. Wadsworth, "Sealice on salmonids: Their biology and control. In Advances in Parasitology", 2000, Vol 44, pp. 233-337

[11] A. Salleb-Aouissi, C. Vrain, and C. Nortet, "QuantMiner: A Genetic Algorithm for Mining Quantitative Association Rules". International Joint Conference on Artificial Intelligence (IJCAI), 2007, pp. 1035-1040

[12] P.-N. Tan, V. Kumar, and J. Srivastava, "Selecting the right objective measure for association analysis", Information Systems, 2004, Vol. 29(4), pp. 293-313

[13] O. Tully, P. Gargan, W. R. Poole, and K. L. Whelan, "Spatial and temporal variation in the infestation of sea trout (Salmo trutta L.) by the caligid copepod Lepeophtheirus salmonis (Kroyer) in relation to sources of infection in Ireland", Parasitology, 1999, Vol. 119, pp. 41-51

[14] G. I. Webb, "Discovering Significant Patterns", Machine Learning, Netherlands: Springer, 2007, Vol. 68(1), pp. 1-33

[15] R. Wille, "Why can concept lattices support knowledge discovery in databases?", Journal of Experimental and Theoretical Artificial Intelligence, 2002, Vol, 14, 2&3, pp. 81–92

Ingunn Alne Hoell has a PhD degree in biochemistry and currently works at Stord/Haugesund University College as an associate professor. Her research interests are within environmental monitoring, marine environment and biotechnology.

Sylvia Encheva, PhD, is professor in mathematics and informatics at Stord/Haugesund University College. Her research interests are within decision support systems, non-classical logics, and fuzzy systems.